

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau**B4**

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification 6 :</b> <b>C07H 17/00, C12N 15/00, 5/00</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 98/20016</b> <b>(43) International Publication Date:</b> 14 May 1998 (14.05.98)
<b>(21) International Application Number:</b> PCT/US97/20016 <b>(22) International Filing Date:</b> 31 October 1997 (31.10.97)  <b>(30) Priority Data:</b> 60/030,279 4 November 1996 (04.11.96) US 60/049,018 9 June 1997 (09.06.97) US  <b>(71) Applicant (for all designated States except US):</b> SMITHKLINE BEECHAM CORPORATION [US/US]; One Franklin Plaza, Philadelphia, PA 19103 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> ESSER, Klaus, Max [US/US]; 200 Hickey Drive, Downingtown, PA 19335 (US). CHAN, John, Y. [US/US]; Apartment 210, 570 West Dekalb Pike, King of Prussia, PA 19406 (US). DABROWSKI-AMARAL, Christine, Ellen [US/US]; 4090 Moscow Road, Parkesburg, PA 19365 (US). DELVECCHIO, Alfred, Michael [US/US]; 668 Metro Court, West Chester, PA 19380 (US). DILLON, Susan, B. [US/US]; 1209 Tullamore Circle, Chester Springs, PA 19425 (US). LEARY, Jeffrey, Joseph [-/US]; 521 General Scott Road, Wayne, PA 19087 (US).		<b>(74) Agents:</b> GEIGER, Kathleen, W. et al.; SmithKline Beecham Corporation, Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US).  <b>(81) Designated States:</b> CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> NOVEL CODING SEQUENCES FROM HERPES SIMPLEX VIRUS TYPE-2  <b>(57) Abstract</b>  This invention relates to newly identified HSV-2 polynucleotides, polypeptides encoded by such polynucleotides, the uses of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides and recombinant host cells transformed with the polynucleotides. This invention also relates to inhibiting or activating the biosynthesis or action of such polynucleotides or polypeptides and the use of such inhibitors or activators in therapy.		

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## NOVEL CODING SEQUENCES FROM HERPES SIMPLEX VIRUS TYPE-2

### Field of the Invention:

This invention relates to newly identified Herpes Simplex Virus type 2 (HSV-2) polynucleotides, polypeptides encoded by such polynucleotides, the uses of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides and recombinant host cells transformed with the polynucleotides. This invention also relates to inhibiting the biosynthesis or action of such polynucleotides or polypeptides and to the use of such inhibitors in therapy of viral infections or related diseases.

### 10 Background of the Invention:

The herpes viruses consist of large icosahedral enveloped virions containing linear double stranded DNA genomes. Currently, eight human herpes viruses have been isolated and are known to be responsible for a variety of disease states, from sub-clinical infections to fatal disease in the immuno compromised. One human herpes virus, herpes simplex virus type 2, designated HSV-2, is usually acquired through sexual contact giving rise to the condition known as genital herpes. The frequency of recurrence of secondary genital herpes ranges between one and six times per year per infected individual. It is estimated that genital HSV-2 infections occur in ten to sixty million individuals in the USA. Less frequently, HSV-2 infection results in herpes labialis, seen as cold sores.

20 General information about HSV-2 may be found in various treatises such as, Herpes Simplex Viruses, In: "Field's Virology", 3rd ed., Lippincott-Raven Publ, pp2297-2342 (1996); Magder, L.S., et al., New England J. Med. 321:7-12 (1989); and "The Human Herpes viruses", Roizman, B. et al., eds. Raven Press, New York, (1993), the contents of which are incorporated herein by reference for purposes of background.

25 Currently, there are no vaccines available to protect against HSV-2 infection. Individuals continue to become infected by the virus and no completely satisfactory anti-viral agents or vaccines are available. Thus HSV-2 presents a major public health problem. There is a need for prophylactic and therapeutic vaccines as well as a method of identifying anti-HSV-2 agents and for reagents useful in such methods. There is a need for a method of identifying compounds which modulate the activity of HSV-2 polynucleotides and proteins and which affect the ability of the virus to replicate and produce multiple infectious virions in an infected cell. There is a need for methods of, and kits for, distinguishing HSV-2 infections from other herpes virus infections.



**Brief Description of the Invention:**

5       Toward these ends, it is an object of the present invention to provide polypeptides, *inter alia*, that have been identified as novel HSV-2 polypeptides by comparison between the amino acid sequences set out in Tables 1-4 and known amino acid sequences of proteins of other viruses such as herpes simplex virus type-1 (HSV-1).

      It is a further object of the invention, to provide polynucleotides that encode HSV-2 proteins, particularly polynucleotides that encode the polypeptides encoded by the Open Reading Frames (ORFs) provided herein, or fragments, analogs or derivatives thereof.

10       In a particularly preferred embodiment of this aspect of the invention the polynucleotides comprise any of the regions encoding HSV-2 proteins in the sequences set out in Tables 1-4, including fragments, analogs or derivatives thereof.

      In another particularly preferred embodiment of the present invention, there is a novel HSV-2 protein comprising any of the amino acid sequences shown in Table 1, or fragments, analogues or derivatives thereof.

15       In accordance with the invention there is provided an isolated nucleic acid molecule encoding a polypeptide expressible by the HSV-2 polynucleotide contained in the deposited HSV-2 strain, SB5.

20       In accordance with the invention there are provided isolated nucleic acid molecules encoding HSV-2 proteins, nucleic acid molecules such as, mRNAs, cDNAs, genomic DNAs and, in further embodiments of this aspect of the invention, biologically, diagnostically, clinically or therapeutically useful variants, analogs or derivatives thereof, or fragments thereof, including fragments of the variants, analogs and derivatives.

      Among the particularly preferred embodiments of this aspect of the invention are naturally occurring allelic variants of HSV-2 proteins.

25       In accordance with this aspect of the invention there are provided novel polypeptides of HSV-2 origin as well as biologically, diagnostically or therapeutically useful fragments thereof, as well as variants, derivatives and analogs of the foregoing and fragments thereof.

30       In accordance with certain preferred embodiments of this and other aspects of the invention there are probes that hybridize to HSV-2 sequences useful for detection of viral infection.

      It also is an object of the invention to provide HSV-2 polypeptides or fragments thereof that may be employed for therapeutic or prophylactic purposes, for example, to treat disease, including treatment by conferring host immunity against viral infections, or as an antiviral agent or a vaccine.

In accordance with another aspect of the present invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization.

Among the particularly preferred embodiments of this aspect of the invention are  
5 variants of HSV-2 polypeptides encoded by naturally occurring alleles of HSV-2 genes for therapeutic or prophylactic use.

It is another object of the invention to provide a process for producing the aforementioned polypeptides, polypeptide fragments, variants and derivatives, fragments of the variants and derivatives, and analogs thereof.

10 In a preferred embodiment of this aspect of the invention there are provided methods for producing the aforementioned HSV-2 polypeptides comprising culturing host cells having expressibly incorporated therein an exogenously-derived HSV-2 encoding polynucleotide under conditions for expression of HSV-2 in the host and then recovering the expressed polypeptide.

In accordance with another object of the invention there are provided products,  
15 compositions, processes and methods that utilize the aforementioned polypeptides and polynucleotides, *inter alia*, for research, biological, clinical, diagnostic, prophylactic and therapeutic purposes.

In accordance with yet another aspect of the present invention, there are provided inhibitors of such polypeptides, useful as antiviral agents. In particular, there are provided  
20 antibodies against such polypeptides. In certain particularly preferred embodiments in this regard, the antibodies are selective for HSV-2.

In a further aspect of the invention there are provided compositions comprising a HSV-2 polynucleotide or HSV-2 polypeptide for administration to cells *in vitro*, to cells *ex vivo* and to cells *in vivo*, or to a multicellular organism. In certain preferred embodiments of this aspect of  
25 the invention, the compositions comprise a HSV-2 polynucleotide for expression of a HSV-2 polypeptide in a host organism to raise an immunological response, preferably to raise immunity in such host against HSV-2 or related organisms.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill from the following description. It should be understood, however, that  
30 the following description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the other parts of the present disclosure.

35

**Detailed Description of the Invention:**

Tables 1-3 show the nucleotide sequences of one strand of "contigs," prepared by assembling sequences derived by sequencing HSV-2, Strain SB5, DNA. Collectively, the contigs herein represent between 85% to over 90% of the genome of this organism. Each of Table 1, 2 and 3 represents a separate sequencing of the HSV-2, SB5, DNA.

Tables 1-3 also show the nucleotide sequences of open reading frames (ORFs), which are deduced DNA coding sequences present within each contig. Tables 1-4 also show the deduced amino acid sequences of polypeptides encoded by these ORFs and sequence homologies to proteins in the NCBI non-redundant protein database.

Each ORF represents a HSV-2 gene although in some cases, a given ORF may actually have been derived from a gene that is longer than the ORF.

Each of the DNA sequences provided herein may be used in the discovery and development of antiviral compounds. For sequences containing an open reading frame (ORF) with appropriate initiation and termination codons, the encoded protein upon expression can be used as a target for the screening of antiviral drugs. Additionally, the DNA sequences encoding preferably the amino terminal regions of the encoded protein, or regions immediately upstream therefrom, can be used to construct antisense sequences to control the expression of the coding sequence of interest. Furthermore, many of the sequences disclosed herein also provide regions upstream and downstream from the encoding sequence. These sequences are useful as a source of regulatory elements for the control of viral gene expression. Such sequences are conveniently isolated by restriction enzyme action or synthesized chemically and introduced, for example, into promoter identification strains. These strains contain a reporter structural gene sequence located downstream from a restriction site such that if an active promoter is inserted, the reporter gene will be expressed.

Although each of the sequences may be employed as described above, this invention also provides several means for identifying particularly useful target genes. The first of these approaches entails searching appropriate databases for sequence matches in related organisms. Thus, if a homologue exists, the HSV-2-like form of this gene would likely play an analogous role. For example, a HSV-2 protein identified as homologous to a cell surface protein in another organism would be useful as a vaccine candidate. To the extent such homologies have been identified for the sequences disclosed herein they are reported along with the encoding sequence in the Tables.

A number of methods can be used to identify genes which are essential to survival *per se*, or essential to the establishment/maintenance of an infection. Identification of an

ORF unknown by one of these methods yields additional information about its function and permits the selection of such an ORF for further development as a screening target. Briefly, these approaches include: generation of temperature sensitive mutations (Weller, S.K., et al., Virology 130:290-305 (1983)), site specific insertion or deletion of a viral gene; a  
5 method based on selection of recombinant molecules generated by double recombination through homologous sequences between intact viral DNA molecules and a DNA fragment containing an insertion or deletion and a selectable marker (Post, L.E., et al., Cell 25:227-32 (1981)), and also by insertional mutagenesis using transposons; a method taking advantage of the random insertion of the DNA phage miniMu into target plasmid DNAs  
10 (Jenkins, F.J., et al., Proc. Nat'l. Acad. Sci. USA 82:4773-4777 (1985)). Each of these techniques may have advantages or disadvantages depending on the particular application. The skilled artisan would choose the approach that is the most relevant with the particular end use in mind. For example, some genes might be recognised as essential for infection but in reality are only necessary for the initiation of infection and so their products would  
15 represent relatively unattractive targets for antivirals developed to cure established and chronic infections.

Use of these technologies when applied to the ORFs of the present invention enables identification of viral proteins expressed during infection, inhibitors of which would have utility in antiviral therapy.

20 **Glossary:**

The following explanations are provided to facilitate understanding of certain terms used frequently herein, particularly in the Examples. The explanations are provided as a convenience and are not limitative of the invention.

HSV-2 BINDING MOLECULE, as used herein, refers to molecules or ions which bind  
25 or interact specifically with HSV-2 polypeptides or polynucleotides of the present invention, including, for example, enzyme substrates, cell membrane components and classical receptors. Binding between polypeptides of the invention and such molecules, including binding or interaction molecules may be exclusive to polypeptides of the invention, which is preferred, or it may be highly specific for polypeptides of the invention, which is also preferred, or it may be  
30 highly specific to a group of proteins that includes polypeptides of the invention, which is preferred, or it may be specific to several groups of proteins at least one of which includes a polypeptide of the invention. Binding molecules also include antibodies and antibody-derived reagents that bind specifically to polypeptides of the invention.

GENETIC ELEMENT generally means a polynucleotide comprising a region that is  
35 important to the viral life cycle, a polynucleotide comprising a region that encodes a polypeptide

or a polynucleotide region that regulates replication, transcription or translation or other processes important to expression of the polypeptide in a host cell, or a polynucleotide comprising both a region that encodes a polypeptide and a region operably linked thereto that regulates expression. Genetic elements may be comprised within a vector that replicates as an episomal element; that is, as a molecule physically independent of the host cell genome. They may be comprised within plasmids. Genetic elements also may be comprised within a host cell genome; not in their natural state but, rather, following manipulation such as isolation, cloning and introduction into a host cell in the form of purified DNA or in a vector, among others.

HOST CELL is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

IDENTITY as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990). As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence it is intended that the nucleotide sequence of the tested polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to

obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. Analogously, by a polypeptide having an amino acid sequence having at least, for example, 95% identity to a reference amino acid sequence is intended that the test amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

ISOLATED means altered "by the hand of man" from its natural state; *i.e.*, that, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a naturally occurring polynucleotide or a polypeptide naturally present in a living organism in its natural state is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. For example, with respect to polynucleotides, the term isolated means that it is separated from the genome and cell in which it naturally occurs. As part of or following isolation, such polynucleotides can be joined to other polynucleotides, such as DNAs, for mutagenesis, to form fusion proteins, and for propagation or expression in a host, for instance. The isolated polynucleotides, alone or joined to other polynucleotides such as vectors, can be introduced into host cells, in culture or in whole organisms. Introduced into host cells in culture or in whole organisms, such DNAs still would be isolated, as the term is used herein, because they would not be in their naturally occurring form or environment. Similarly, the polynucleotides and polypeptides may occur in a composition, such as a media formulations, solutions for introduction of polynucleotides or polypeptides, for example, into cells, compositions or

solutions for chemical or enzymatic reactions, for instance, which are not naturally occurring compositions, and, therein remain isolated polynucleotides or polypeptides within the meaning of that term as it is employed herein.

POLYNUCLEOTIDE(S) generally refers to any polyribonucleotide or  
5 polydeoxiribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. Thus, for instance, polynucleotides as used herein refers to, among others, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be  
10 single-stranded or, more typically, double-stranded, or triple-stranded, or a mixture of single- and double-stranded regions. In addition, polynucleotide as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the  
15 molecules of a triple-helical region often is an oligonucleotide. As used herein, the term polynucleotide includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two  
20 examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including *inter alia*,  
25 simple and complex cells. The term polynucleotide(s) embrace short polynucleotides often referred as oligonucleotides.

POLYPEPTIDES, as used herein, includes all polypeptides as described below. The basic structure of polypeptides is well known and has been described in innumerable textbooks and other publications in the art. In this context, the term is used herein to refer to any peptide or  
30 protein comprising two or more amino acids joined to each other in a linear chain by peptide bonds. As used herein, the term refers to both short chains, which also commonly are referred to in the art as peptides, oligopeptides and oligomers, for example, and to longer chains, which generally are referred to in the art as proteins, of which there are many types. It will be appreciated that polypeptides often contain amino acids other than the 20 amino acids commonly  
35 referred to as the 20 naturally occurring amino acids, and that many amino acids, including the

terminal amino acids, may be modified in a given polypeptide, either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques which are well known to the art. Even the common modifications that occur naturally in polypeptides are too numerous to list exhaustively here, but they are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. Among the known modifications which may be present in polypeptides of the present are, to name an illustrative few, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Such modifications are well known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, F., *Posttranslational Protein Modifications: Perspectives and Prospects*, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter *et al.*, *Meth. Enzymol.* 182:626-646 (1990) and Rattan *et al.*, *Protein Synthesis: Posttranslational Modifications and Aging*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992). It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in



polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli* or other cells, prior to proteolytic processing, almost invariably will be N-formylmethionine. During post-translational modification of the peptide, a methionine residue at the NH<sub>2</sub>-terminus may be deleted. Accordingly, this invention contemplates the use of both the methionine-containing and the methionineless amino terminal variants of the protein of the invention. The modifications that occur in a polypeptide often will be a function of how it is made. For polypeptides made by expressing a cloned gene in a host, for instance, the nature and extent of the modifications in large part will be determined by the host cell posttranslational modification capacity and the modification signals present in the polypeptide amino acid sequence. For instance, as is well known, glycosylation often does not occur in bacterial hosts such as, for example, *E. coli*. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the same posttranslational glycosylations as do mammalian cells and, for this reason, insect cell expression systems have been developed to express efficiently mammalian proteins having native patterns of glycosylation. Similar considerations apply to other modifications. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. In general, as used herein, the term polypeptide encompasses all such modifications, particularly those that are present in polypeptides synthesized by expressing a polynucleotide in a host cell.

VARIANT(S) as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of

polynucleotides and polypeptides may be made by mutagenesis techniques, by direct synthesis, and by other recombinant methods known to skilled artisans.

**Deposit:**

HSV-2, strain SB5 has been deposited at the American Type Culture Collection  
5 under accession number VR-2546 on October 31, 1996.

The deposits referred to herein will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. These deposits are provided merely as convenience to those of skill in the art and are not an admission that a deposit is required under 35 U.S.C. §112. The  
10 sequence of the polynucleotides contained in the deposited material, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with any description of sequences herein. A license may be required to make, use or sell the deposited material, and no such license is hereby granted.

15 **Viral Strain and Genome:**

The nucleotide sequences disclosed herein can be obtained by synthetic chemical techniques known in the art or can be obtained from HSV-2, strain SB5 by probing a DNA preparation with probes constructed from the particular sequences disclosed herein. Alternatively, oligonucleotides derived from a disclosed sequence can act as PCR primers  
20 in a process of PCR-based cloning of the sequence from a viral genomic source. It is recognised that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The present invention relates to novel HSV-2 polypeptides and polynucleotides encoding same, among other things, as described in greater detail below. The invention relates  
25 especially to HSV-2 molecules having the nucleotide and amino acid sequences set out in Tables 1-4 and to the HSV-2 nucleotide and amino acid sequences of the DNA isolatable from Deposit No. ATTC VR-2546, which is herein referred to as "the deposited organism" or as the "DNA of the deposited organism." It will be appreciated that the nucleotide and amino acid sequences set out in Tabled 1-4 were obtained by sequencing the DNA of the deposited organism. Hence, the  
30 sequence of the deposited clone is controlling as to any discrepancies between it (and the sequence it encodes) and the sequences of the Tables.

The present invention also relates to additional polynucleotide sequences disclosed herein, which are RNAs transcribed from the DNAs disclosed herein but which may or may not be translated into protein. Such polynucleotides are known in HSV-1 and other herpes viruses.

### Polynucleotides

In accordance with one aspect of the present invention, there are provided isolated polynucleotides which encode HSV-2 polypeptides having the deduced amino acid sequence of Tables 1-4. It is preferred that these polynucleotides be one of those set forth in Tables 1, 2 or 3.

- 5 The skilled artisan can readily determine the polynucleotide sequence of such preferred polynucleotides by reference to the ORF start and stop positions set forth in Tables 1-4.

Using the information provided herein, such as the polynucleotide sequence set out in Tables 1-3, a polynucleotide of the present invention encoding HSV-2 polypeptide may be obtained using standard cloning and screening procedures. To obtain the polynucleotide  
10 encoding the protein using the DNA sequences given in Tables 1-3 typically a library of clones of chromosomal DNA of HSV-2 strain SB5 in *E. coli* or some other suitable host is probed with a radiolabelled oligonucleotide, preferably a 17mer or longer, derived from a sequence of Tables 1-3. Clones carrying DNA identical to that of the probe can then be distinguished using high stringency washes. By sequencing the individual clones thus identified with sequencing primers  
15 designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook, J. in MOLECULAR CLONING, A Laboratory Manual (2nd edition 1989 Cold Spring Harbor Laboratory. see Screening By  
20 Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70

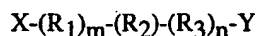
The DNA sequences set out in Tables 1, 2 and 3 each contain at least one open reading frame encoding a protein having at least about the number of amino acid residues set forth in Table 1-3. The start and stop codons of each open reading frame are the first three and the last three nucleotides of each polynucleotide set forth in Table 1, 2 and 3.

- 25 Certain HSV-2 sequences of the invention are structurally related to sequences encoding other proteins of the herpes family, as shown by comparing the sequences of the Tables with that of sequences reported in the literature. Moreover, certain polynucleotides and polypeptides of the invention are structurally related to known. These proteins exhibit greatest homology to the homologue listed in Tables 1, 2, 3 and 4 from among the known proteins.

- 30 The invention provides a polynucleotide sequence identical over its entire length to each coding sequence in Tables 1-3. Also provided by the invention is the coding sequence for the mature polypeptide or a fragment thereof, by itself as well as the coding sequence for the mature polypeptide or a fragment in reading frame with other coding sequence, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence. The polynucleotide  
35 may also contain non-coding sequences, including for example, but not limited to non-coding 5'

and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain  
 5 embodiments of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA* 86: 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell* 37: 767 (1984). Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

10 The invention also includes polynucleotides of the formula:



wherein, at the 5' end of the molecule, X is hydrogen, and at the 3' end of the molecule, Y is hydrogen or a metal,  $R_1$  and  $R_3$  is any nucleic acid residue, n and/or m is an integer between 1 and 3000 or zero, and  $R_2$  is a nucleic acid sequence of the invention, particularly a nucleic acid  
 15 sequence selected from the group set forth in Tables 1, 2 and 3, as well as a ORF sequence selected from the group set forth in Tables 1, 2, 3 and 4 (as indicated by the reading frame numbering). In the polynucleotide formula above  $R_2$  is oriented so that its 5' end residue is at the left, bound to  $R_1$ , and its 3' end residue is at the right, bound to  $R_3$ . Any stretch of nucleic acid residues denoted by either R group, where n and/or m is greater than 1, may be either a  
 20 heteropolymer or a homopolymer, preferably a heteropolymer. In a preferred embodiment n and/or m is an integer between 1 and 1000, or 2000 or 3000.

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a viral polypeptide and more particularly a polypeptide of the HSV-2 having an amino acid  
 25 sequence set out in Table 1, 2, 3 or 4. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or an insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that  
 30 encode for variants of the polypeptide having the deduced amino acid sequence of Tables 1, 2, 3 and 4. Variants that are fragments of the polynucleotides of the invention may be used to synthesize full-length polynucleotides of the invention.

Polynucleotides of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or  
 35 produced by chemical synthetic techniques or by a combination thereof. The DNA may be

double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The coding sequence which encodes the polypeptide may be identical to the coding sequence of the polynucleotide shown in Tables 1-4. It also may be a polynucleotide with a  
5 different sequence, which, as a result of the redundancy (degeneracy) of the genetic code, encodes the polypeptides of Tables 1-4.

Particularly preferred embodiments are polynucleotides encoding polypeptide variants, that have the amino acid sequence of a polypeptide of Tables 1, 2, 3 and/or 4 in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in  
10 any combination. Especially preferred among these are silent substitutions, additions and deletions, that do not alter the properties and activities of such polynucleotide.

Further preferred embodiments of the invention are polynucleotides that are at least 50%, 60% or 70% identical over their entire length to a polynucleotide encoding a polypeptide having the amino acid sequence set out in Tables 1, 2, 3 or 4, and polynucleotides that are  
15 complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding a polypeptide of the deposited strain and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length to the same are particularly preferred, and among these particularly preferred  
20 polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the most preferred.

A preferred embodiment is an isolated polynucleotide comprising a polynucleotide  
25 sequence selected from the group consisting of: a polynucleotide having at least a 50% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Tables 1, 2, 3, or 4 and obtained from a prokaryotic species other than HSV-2; and a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 50% identical to the amino acid sequence of Tables 1, 2, 3 or 4 and obtained from a prokaryotic species other than HSV-2.

30 Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptide encoded by the DNA of Tables 1, 2, 3 or 4.

The invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the invention especially relates to polynucleotides that hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C. Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in Tables 1, 2, 3 or 4 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or a fragment thereof; and isolating said DNA sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers described elsewhere herein.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding a polypeptide and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to a polynucleotide set forth in Table 1, 2, 3 or 4. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of each gene that comprises or is comprised by a polynucleotide set forth in Table 1, 2, 3 or 4 may be isolated by screening using a DNA sequence provided in Table 1, 2, 3 or 4 to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

Polynucleotides of the invention that are oligonucleotides derived from the a polynucleotide or polypeptide sequence set forth in Table 1, 2, 3 or 4 may be used in the processes herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are transcribed in virus in infected  
5 tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The invention also provides polynucleotides that may encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for  
10 instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

15 A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

The DNA may also comprise a promoter region which functions to direct the  
20 transcription of the mRNA encoding the HSV-2 of this invention. Such promoters may be independently useful to direct the transcription of heterologous genes in recombinant expression systems. Polyadenylation and splicing signal sequences are also present in the polynucleotide sequence and may be useful as gene expression signal in heterologous gene expression vectors and constructs.

25 The polynucleotides and polypeptides of the invention may be employed, for example, as research reagents and materials for discovery of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to polynucleotide assays.

The polynucleotides of the invention that are oligonucleotides may also be used as nucleic acid amplification primers, such as PCR primers, in the process herein described to  
30 determine whether or not the HSV-2 genes identified herein in whole or in part are present or transcribed in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

In addition to the uses mentioned above for the polynucleotides of this invention, the following applications are also contemplated by this invention. *Inter alia*, the  
35 polynucleotides disclosed herein or portions thereof, may be used as probes to discover

- mRNA transcripts synthesized during productive and latent HSV-2 infections, for example by Northern blot, nuclease protection, and primer extension experiments. Novel transcripts in turn can lead to the discovery of new HSV-2 proteins not deducible from the genome sequences directly. The sequences, or portions thereof, may be used to discover antisense inhibitors of virus replication and novel therapeutics based on antisense mechanisms. The sequences, or portions thereof, may be used to prepare novel gene therapy vectors. The sequences or portions thereof may be used as a basis for the generation of DNA- or RNA-containing oligonucleotides designed to form a triplex with duplex DNA, for use as analytical tools, diagnostics or therapeutics.
- 5 Nucleic acid sequences, or portion thereof, can be used to generate cell lines useful for diagnostics or screening. The DNA sequences can be used to predict restriction enzyme sites useful for replacing the gene in the viral genome with a marker gene such as *lac z* or green fluorescent protein. Such a replacement is useful in defining the biological role of the gene in the viral life cycle. These gene knockout experiments are useful to discover genes which are likely to be high quality drug discovery targets (essential genes) or good locations for foreign genes for the purposes of gene therapy (non-essential genes) through an HSV-2 viral vector. Such gene replacements are also useful for discovering virulence factors, for example by comparing the pathogenicity of the modified virus with the unmodified virus or through the ease of identifying a marker gene such as *lac z*.
- 10 15 20

In addition to the standard A, G, C, T/U representations for nucleic acid bases, the term "N" is also used. "N" means that any of the four DNA or RNA bases may appear at such a designated position in the DNA or RNA sequence, except it is preferred that N is not a base that when taken in combination with adjacent nucleotide positions, when read in the correct reading frame, would have the effect of generating a premature termination codon in such reading frame.

25

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

30

#### Polypeptides

The present invention further relates to HSV-2 polypeptides that have the deduced amino acid sequences of the polypeptides defined by amino acid sequence in Tables 1-4.

35



The invention also relates to fragments, analogs and derivatives of these polypeptides. The terms "fragment," "derivative" and "analog" when referring to the polypeptides of the invention mean a polypeptide which retains essentially the same biological function or activity as such polypeptide. Fragments, derivatives and analogs that retain at least 90% of the biological activity of the native HSV-2 protein are preferred. Fragments, derivatives and analogs that retain at least 95% of the activity of the native HSV-2 protein are preferred. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide.

The polypeptide of the present invention may be a recombinant polypeptide, a natural polypeptide or a synthetic polypeptide. In certain preferred embodiments it is a recombinant polypeptide.

The fragment, derivative or analog of the polypeptides of the invention may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be obtained by those of ordinary skill in the art, from the teachings herein.

Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Further particularly preferred in this regard are variants, analogs, derivatives and fragments having the amino acid sequence of one or more of the HSV-2 polypeptides of the invention, in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the HSV-

2 protein. Also especially preferred in this regard are conservative substitutions. Most highly preferred are polypeptides having the amino acid sequences of Tables 1-4 without substitutions.

The invention also includes polypeptides of the formula:



5 wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal,  $R_1$  and  $R_3$  are any amino acid residue, n and/or m is an integer between 1 and 2000 or zero, and  $R_2$  is an amino acid sequence of the invention, particularly an amino acid sequence selected from the group set forth in Tables 1, 2, 3 and 4. In the formula above  $R_2$  is oriented so that its amino terminal residue is at the left, bound to  $R_1$ , and its carboxy terminal residue is at  
10 the right, bound to  $R_3$ . Any stretch of amino acid residues denoted by either R group, where n and/or m is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. In preferred embodiments n and/or m is an integer between 1 and 1000 or 2000.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

15 The polypeptides of the present invention include the polypeptides of Tables 1-4, in particular the mature polypeptide as well as polypeptides which have at least 60%, 70% or 80% identity to one or more of the polypeptides of Tables 1-4 and preferably at least 90% similarity to one or more of the polypeptides of Tables 1-4 and more preferably at least 95% similarity; and  
20 also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 contiguous amino acids and more preferably at least 50 contiguous amino acids.

In addition to the uses mentioned above for the polypeptides of this invention, the following applications are also contemplated by this invention. *Inter alia*, the polypeptides  
25 disclosed herein or portions thereof which have enzymatic activity or structural functionality are useful as a source of those proteins for screening and or therapy. Such polypeptides may be identified by homology for example, to HSV1 polypeptides that code for proteins with known function (e.g., helicases, kinases, proteases). Use of polypeptides of the invention for screening or therapy based upon functionality predicted by homology match is  
30 a particularly preferred aspect of this invention. Also the polypeptides derived from the deposited strain ATCC VR-2546 herein can be used for comparison with sequences from other HSV-2 strains in the public domain, for example, comparison of the polypeptides of the invention with strain HG52 may be useful in the discovery of virulence factors, since HG52 is avirulent in mouse and guinea pig infection models and HSV-2 SB5 is virulent.  
35 Similarly, public domain homolog from strain MS may be useful in the discovery of

virulence factors since there are major differences in the CNS pathogenesis in animal models between strains MS and SB5.

In addition to the standard single and triple letter representations for amino acids, the term "X" or "Xaa" is also used. "X" and "Xaa" mean that any of the twenty naturally occurring amino acids may appear at such a designated position in the polypeptide sequence.

### Fragments

Fragments or portions of the polypeptides of the present invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides.

Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

Also among preferred embodiments of this aspect of the present invention are polypeptides comprising fragments of HSV-2, most particularly fragments of HSV-2 having the amino acid sequences set out in Tables 1-4, and variants and derivatives thereof.

In this regard, a fragment is a polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned HSV-2 polypeptides and variants or derivatives thereof.

Such fragments may be "free-standing," *i.e.*, not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the presently discussed fragments most preferably form a single continuous region. However, several fragments may be comprised within a single larger polypeptide. For instance, certain preferred embodiments relate to a fragment of a HSV-2 polypeptide of the present comprised within a precursor polypeptide designed for expression in a host and having heterologous pre and pro-polypeptide regions fused to the amino terminus of the HSV-2 fragment and an additional region fused to the carboxyl terminus of the fragment. Therefore, fragments in one aspect of the meaning intended herein, refers to the portion or portions of a fusion polypeptide or fusion protein derived from HSV-2.

Representative examples of polypeptide fragments of the invention, include, for example, those which have from about 5-15, 10-20, 15-40, 30-55, 41-75, 41-80, 41-90, 50-100, 75-100, 90-115, 100-125, and 110-140, 120-150, 200-300, 1-175, 1-600 or 1-1000 amino acids long. Particular examples of polypeptide fragments of the inventions that may be mentioned include fragments of 20-200 amino acids.

In this context about includes the particularly recited range and ranges larger or smaller by several, a few, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes.

Among especially preferred fragments of the invention are truncation mutants of HSV-2. Truncation mutants include HSV-2 polypeptides having the amino acid sequences of Tables 1-4, or of variants or derivatives thereof, except for deletion of a continuous series of residues (that is, a continuous region, part or portion) that includes the amino terminus, or a continuous series of  
5 residues that includes the carboxyl terminus or, as in double truncation mutants, deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Fragments having the size ranges set out above also are preferred embodiments of truncation fragments, which are especially preferred among fragments generally. Degradation forms of the polypeptides of the invention in a host cell are also preferred.

10 Also preferred in this aspect of the invention are fragments characterized by structural or functional attributes of HSV-2. Preferred embodiments of the invention in this regard include fragments that comprise alpha-helix and alpha-helix forming regions ("alpha-regions"), beta-sheet and beta-sheet-forming regions ("beta-regions"), turn and turn-forming regions ("turn-regions"), coil and coil-forming regions ("coil-regions"), hydrophilic regions, hydrophobic  
15 regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions and high antigenic index regions of HSV-2.

Further preferred regions are those that mediate activities of HSV-2. Most highly preferred in this regard are fragments that have a chemical, biological or other activity of the particular HSV-2 protein, including those with a similar activity or an improved activity, or with  
20 a decreased undesirable activity. Routinely one generates the fragment by well-known methods then compares the activity of the fragment to the native protein in a convenient assay such as listed hereinbelow. Highly preferred in this regard are fragments that contain regions that are homologs in sequence, or in position, or in both sequence and to active regions of related polypeptides, such as the related polypeptides set out in Table 1. Among particularly preferred  
25 fragments in these regards are truncation mutants, as discussed above. Further preferred polynucleotide fragments are those that are antigenic or immunogenic in an animal, especially in a human.

It will be appreciated that the invention also relates to, among others, polynucleotides encoding the aforementioned fragments, polynucleotides that hybridize to polynucleotides  
30 encoding the fragments, particularly those that hybridize under stringent conditions, and polynucleotides, such as PCR primers, for amplifying polynucleotides that encode the fragments. In these regards, preferred polynucleotides are those that correspond to the preferred fragments, as discussed above.

Vectors, host cells, expression:

The present invention also relates to vectors which comprise a polynucleotide or polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

5 Host cells can be genetically engineered to incorporate polynucleotides and express polypeptides of the present invention. Introduction of a polynucleotides into the host cell can be affected by calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction, infection or other methods. Such methods are described in many standard  
10 laboratory manuals, such as Davis *et al.*, *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook *et al.*, *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

Polynucleotide constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the  
15 invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are  
20 described by Sambrook *et al.*, *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

In accordance with this aspect of the invention the vector may be, for example, a plasmid vector, a single or double-stranded phage vector, a single or double-stranded RNA or DNA viral vector. Plasmids generally are designated herein by a lower case p preceded and/or  
25 followed by capital letters and/or numbers, in accordance with standard naming conventions that are familiar to those of skill in the art. Starting plasmids disclosed herein are either commercially available, publicly available, or can be constructed from available plasmids by routine application of well known, published procedures. Many plasmids and other cloning and expression vectors that can be used in accordance with the present invention are well known and  
30 readily available to those of skill in the art.

Preferred among vectors, in certain respects, are those for expression of polynucleotides and polypeptides of the present invention. Generally, such vectors comprise *cis*-acting control regions effective for expression in a host operatively linked to the polynucleotide to be expressed. Appropriate *trans*-acting factors either are supplied by the host, supplied by a complementing  
35 vector or supplied by the vector itself upon introduction into the host.

In certain preferred embodiments in this regard, the vectors provide for specific expression. Such specific expression may be inducible expression or expression only in certain types of cells or both inducible and cell-specific. Particularly preferred among inducible vectors are vectors that can be induced for expression by environmental factors that are easy to  
5 manipulate, such as temperature and nutrient additives. A variety of vectors suitable to this aspect of the invention, including constitutive and inducible expression vectors for use in prokaryotic and eukaryotic hosts, are well known and employed routinely by those of skill in the art.

A great variety of expression vectors can be used to express a polypeptide of the  
10 invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, *e.g.*, vectors derived from viral plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as  
15 those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids, all may be used for expression in accordance with this aspect of the present invention. Generally, any vector suitable to maintain, propagate or express polynucleotides to express a polypeptide in a host may be used for expression in this regard.

The appropriate DNA sequence may be inserted into the vector by any of a variety of  
20 well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989).

The DNA sequence in the expression vector is operatively linked to appropriate expression control sequence(s), including, for instance, a promoter to direct mRNA transcription.  
25 Representatives of such promoters include, but are not limited to, the phage lambda PL promoter, the *E. coli* lac, trp and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs.

In general, expression constructs will contain sites for transcription initiation and termination, and, in some instances, in the transcribed region, a ribosome binding site for  
30 translation. The coding portion of the mature transcripts expressed by the constructs will include a translation initiating AUG at the beginning and a termination codon appropriately positioned at the end of the polypeptide to be translated.

In addition, the constructs may contain control regions that regulate as well as engender expression. Generally, in accordance with many commonly practiced procedures, such regions

will operate by controlling transcription, such as transcription factors, repressor binding sites and termination, among others.

Vectors for propagation and expression generally will include selectable markers and amplification regions, such as; for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989).

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

The following vectors, which are commercially available, are provided by way of example. Among vectors preferred for use in bacteria are pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; and ptc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia, and pBR322 (ATCC 37017). Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. These vectors are listed solely by way of illustration of the many commercially available and well known vectors that are available to those of skill in the art for use in accordance with this aspect of the present invention. It will be appreciated that any other plasmid or vector suitable for, for example, introduction, maintenance, propagation or expression of a polynucleotide or polypeptide of the invention in a host may be used in this aspect of the invention.

Promoter regions can be selected from any desired gene using vectors that contain a reporter transcription unit lacking a promoter region, such as a chloramphenicol acetyl transferase ("CAT") transcription unit, downstream of restriction site or sites for introducing a candidate promoter fragment; *i.e.*, a fragment that may contain a promoter. As is well known, introduction into the vector of a promoter-containing fragment at the restriction site upstream of the *cat* gene engenders production of CAT activity, which can be detected by standard CAT assays. Vectors suitable to this end are well known and readily available, such as pKK232-8 and pCM7.

Promoters for expression of polynucleotides of the present invention include not only well known and readily available promoters, but also promoters that readily may be obtained by the foregoing technique, using a reporter gene.

Among known prokaryotic promoters suitable for expression of polynucleotides and polypeptides in accordance with the present invention are the *E. coli* *lacI* and *lacZ* and

promoters, the T3 and T7 promoters, the *gpt* promoter, the lambda PR, PL promoters and the *trp* promoter.

Among known eukaryotic promoters suitable in this regard are the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the  
5 promoters of retroviral LTRs, such as those of the Rous sarcoma virus ("RSV"), and metallothionein promoters, such as the mouse metallothionein-I promoter.

Recombinant expression vectors will include, for example, origins of replication, a promoter preferably derived from a highly-expressed or regulatable gene to direct transcription of a downstream structural sequence, and a selectable marker to permit isolation of vector  
10 containing cells after exposure to the vector.

Polynucleotides of the invention, encoding the heterologous structural sequence of a polypeptide of the invention generally will be inserted into the vector using standard techniques so that it is operably linked to the promoter for expression. The polynucleotide will be positioned so that the transcription start site is located appropriately 5' to the AUG that initiates  
15 translation of the polypeptide to be expressed. Where applicable, a ribosome binding site may be located between the transcription start site and the initiating AUG. Generally, there will be no other open reading frames that begin with an initiation codon, usually AUG, and lie between the ribosome binding site, where applicable or the 5' end of the transcript and the initiation codon. Also, generally, there will be a translation stop codon at the end of the polypeptide and there will  
20 be a polyadenylation signal in constructs for use in eukaryotic hosts. Transcription termination signal appropriately disposed at the 3' end of the transcribed region may also be included in the polynucleotide construct.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be  
25 incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to  
30 the N-or C-terminus of the polypeptide to improve stability and persistence in the host cell, during purification or during subsequent handling and storage. Also, a region may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability or to facilitate purification, among others, are familiar and routine  
35 techniques in the art. A preferred fusion protein comprises a heterologous region from



immunoglobulin that is useful to solubilize or purify polypeptides. For example, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another protein or part thereof. In drug discovery, for example, proteins have been fused with antibody Fc portions  
5 for the purpose of high-throughput screening assays to identify antagonists. See, D. Bennett et al., Journal of Molecular Recognition, 8: 52-58 (1995) and K. Johanson et al., The Journal of Biological Chemistry, 270,(16): 9459-9471 (1995).

Cells typically then are harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

10 Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well know to those skilled in the art.

Mammalian expression vectors may comprise an origin of replication, a suitable promoter and enhancer, and also any necessary polyadenylation sites, splice donor and acceptor  
15 sites, transcriptional termination sequences, and 5' flanking non-transcribed sequences that are necessary for expression. In certain preferred embodiments in this regard DNA sequences derived from the SV40 splice sites, and the SV40 polyadenylation sites are used for required non-transcribed genetic elements of these types.

HSV-2 polypeptides can be recovered and purified from recombinant cell cultures by  
20 well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification. Well known techniques for refolding protein may be employed to  
25 regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, viral, yeast, higher plant, insect and  
30 mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

HSV-2 polynucleotides and polypeptides may be used in accordance with the present  
35 invention for a variety of applications, particularly those that make use of the chemical and

biological properties of HSV-2. Additional applications relate to diagnosis and to treatment of disorders of cells, tissues and organisms. These aspects of the invention are illustrated further by the following discussion.

**Polynucleotide assays:**

- 5           This invention is also related to the use of the HSV-2 polynucleotides to detect complementary polynucleotides such as, for example, as a diagnostic reagent. Detection of HSV-2 polynucleotides in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method that can add to, define or allow a diagnosis of a disease. Eukaryotes (herein also "individual(s)"), particularly mammals, and especially humans, infected by HSV-2 may be
- 10 detected at the DNA or RNA level by a variety of techniques. Nucleic acids for diagnosis may be obtained from an individual's cells, tissues, and fluids, such as brain, bone, blood, muscle, cartilage, skin, saliva, urine, semen, and mucous. Tissue biopsy and autopsy material is also preferred for samples from an individual to use in a diagnostic assay. The viral DNA may be used directly for detection or may be amplified enzymatically by using PCR prior to analysis
- 15 (Saiki *et al.*, *Nature* 324: 163-166 (1986)). RNA or cDNA may also be used in the same ways. As an example, PCR primers complementary to the nucleic acid encoding HSV-2 can be used to identify and analyze HSV-2 presence and expression. Using PCR, characterization of the strain of virus present in a eukaryote, particularly a mammal, and especially a human, may be made by an analysis of the genotype of the viral gene. For example, deletions and insertions can be
- 20 detected by a change in size of the amplified product in comparison to the genotype of a reference sequence. Point mutations can be identified by hybridizing amplified DNA to radiolabeled HSV-2 RNA or alternatively, radiolabeled HSV-2 antisense DNA sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures.
- 25           Sequence differences between a reference gene and genes having mutations also may be revealed by direct DNA sequencing. In addition, cloned DNA segments may be employed as probes to detect specific DNA segments. The sensitivity of such methods can be greatly enhanced by appropriate use of PCR or another amplification method. For example, a sequencing primer is used with double-stranded PCR product or a single-stranded template
- 30 molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabeled nucleotide or by automatic sequencing procedures with fluorescent-tags.

- Genetic typing of various strains of virus based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments in gels, with or
- 35 without denaturing agents. Small sequence deletions and insertions can be visualized by high

resolution gel electrophoresis. DNA fragments of different sequences may be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial melting temperatures (see, *e.g.*, Myers *et al.*, Science, 230: 1242 (1985)).

5       Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (*e.g.*, Cotton *et al.*, Proc. Nat'l. Acad. Sci., USA, 85: 4397-4401 (1985)).

10       Thus, the detection of a specific DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing or the use of restriction enzymes, (*e.g.*, restriction fragment length polymorphisms ("RFLP") and Southern blotting of genomic DNA.

In addition to more conventional gel-electrophoresis and DNA sequencing, mutations also can be detected by *in situ* analysis.

15       Cells carrying mutations or polymorphisms in the gene of the present invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example. Nucleic acids for diagnosis may be obtained from an infected individual's cells, including but not limited to blood, urine, saliva, tissue biopsy and autopsy material or from virus isolated and cultured from the above or other sources. The viral DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki *et al.*, Nature, 324:163-166  
20       (1986)) prior to analysis. RT-PCR can also be used to detect mutations. It is particularly preferred to used RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA or cDNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to the nucleic acid encoding HSV-2 can be used to identify and analyze mutations. For example, deletions and insertions can be detected by a  
25       change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabeled RNA or alternatively, radiolabeled antisense DNA sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures. The primers may be used to amplify the gene isolated from the individual such that the gene may  
30       then be subject to various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be detected.

#### Polypeptide assays:

The present invention also relates to diagnostic assays such as quantitative and diagnostic assays for detecting levels of HSV-2 protein in cells and tissues, including  
35       determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in

accordance with the invention for detecting expression of HSV-2 protein compared to normal control tissue samples may be used to detect the presence of an infection. Assay techniques that can be used to determine levels of a protein, such as an HSV-2 protein of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays. Among these ELISAs frequently are preferred. An ELISA assay initially comprises preparing an antibody specific to HSV-2, preferably a monoclonal antibody. In addition a reporter antibody generally is prepared which binds to the monoclonal antibody. The reporter antibody is attached to a detectable reagent such as radioactive, fluorescent or enzymatic reagent, in this example horseradish peroxidase enzyme.

#### Antibodies:

The polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. The present invention includes, for examples monoclonal and polyclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library.

Antibodies generated against the polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.

For preparation of monoclonal antibodies, any technique known in the art which provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies to immunogenic polypeptide products of this invention.

Alternatively phage display technology could be utilized to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-Fbp or from naive libraries

(McCafferty, J. *et al.*, Nature 348, 552-554 (1990); Marks, J. *et al.*, *Biotechnology* 10: 779-783 (1992). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, *Nature* 352, 624-628 (1991).

5 If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or purify the polypeptide of the present invention by attachment of the antibody to a solid support for isolation and/or purification by affinity chromatography.

10 Thus among others, antibodies against HSV-2 may be employed to inhibit and/or treat infections, particularly viral infections, and especially HSV-2 infections as well as to monitor the effectiveness of antibiotic treatment.

Polypeptide derivatives include antigenically, epitopically or immunologically equivalent derivatives which form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its  
15 equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or polypeptide according to the present invention, interfere with the immediate physical interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the  
20 antibodies act to interfere with the immediate physical interaction between pathogen and mammalian host.

The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof, is used as an antigen to immunize a mouse or other animal such as a rabbit, rat or chicken. The fusion protein may provide stability to the polypeptide. The  
25 antigen may be associated, for example by conjugation, with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively a multiple antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

30 Preferably the antibody or derivative thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanised"; where the complementarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for example as described in Jones, P. *et al.* *Nature* 321: 522-525 (1986) or Tempest *et al.*,  
35 *Biotechnology* 9: 266-273 (1991). The above antibody reagents will also be useful for

assessing the biological role of the gene through antibody inhibition studies, immunoprecipitation studies, super-shift experiments and similar techniques. These studies may lead to discovery of novel protein:protein interactions which may be useful drug targets. The above antibody reagents may lead to the identification of novel viral proteins not predicted by the DNA sequence, which in turn may be novel drug targets.

**HSV-2 binding molecules and assays:**

This invention also provides a method for identification of molecules, such as binding molecules, that bind HSV-2. Genes encoding proteins that bind HSV-2, such as binding proteins, can be identified by numerous methods known to those of skill in the art, for example, ligand panning and FACS sorting. Such methods are described in many laboratory manuals such as, for instance, Coligan *et al.*, *Current Protocols in Immunology* 1(2): Chapter 5 (1991).

For instance, expression cloning may be employed for this purpose. To this end polyadenylated RNA is prepared from a cell expressing HSV-2, a cDNA library is created from this RNA, the library is divided into pools and the pools are transfected individually into cells that are not expressing HSV-2. The transfected cells then are exposed to labeled HSV-2. HSV-2 can be labeled by a variety of well-known techniques including standard methods of radioiodination or inclusion of a recognition site for a site-specific protein kinase. Following exposure, the cells are fixed and binding of HSV-2 is determined. These procedures conveniently are carried out on glass slides.

Alternatively a labeled ligand can be photoaffinity linked to a cell extract, such as a membrane or a membrane extract, prepared from cells that express a molecule that it binds, such as a binding molecule. Cross-linked material is resolved by polyacrylamide gel electrophoresis ("PAGE") and exposed to X-ray film. The labeled complex containing the ligand-binding can be excised, resolved into peptide fragments, and subjected to protein microsequencing. The amino acid sequence obtained from microsequencing can be used to design unique or degenerate oligonucleotide probes to screen cDNA libraries to identify genes encoding the putative binding molecule.

Polypeptides of the invention also can be used to assess HSV-2 binding capacity of HSV-2 binding molecules in cells or in cell-free preparations.

Polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics.

This invention also provides a method of screening drugs to identify those which interfere with the proteins selected as targets herein, which method comprises measuring the

interference of the activity of the protein by a test drug. For example if the protein selected has a catalytic activity, after suitable purification and formulation the activity of the enzyme can be followed by its ability to convert its natural substrates. By incorporating different chemically synthesised test compounds or natural products into such an assay of enzymatic activity one is able to detect those additives which compete with the natural substrate or otherwise inhibit enzymatic activity.

The invention also relates to activators and inhibitors identified thereby.

Another aspect of the invention relates to use of a polynucleotide in genetic immunization, and will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff *et al.*, Hum. Mol. Genet. 1:363 (1992); Manthorpe *et al.*, Hum. Gene Ther. 4:419 (1993)), delivery of DNA complexed with specific protein carriers (Wu *et al.*, J. Biol. Chem. 264:16985 (1989)), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, Proc. Nat'l Acad. Sci. USA, 83:9551 (1986)), encapsulation of DNA in various forms of liposomes (Kaneda *et al.*, Science 243:375 (1989)), particle bombardment (Tang *et al.*, Nature 356:152 (1992)); Eisenbraun *et al.*, DNA Cell Biol. 12:791 (1993)) and *in vivo* infection using cloned retroviral vectors (Seeger *et al.*, Proc. Nat'l. Acad. Sci. USA 81:5849 (1984)). Suitable promoters for muscle transfection include CMV, RSV, SRA, actin, MCK, alpha globin, adenovirus and dihydrofolate reductase.

In therapy or as a prophylactic, the active agent i.e., the polypeptide, polynucleotide or inhibitor of the invention, may be administered to a patient as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

#### Vaccines:

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal which comprises inoculating the individual with HSV-2 polypeptide, or an antigenic fragment or variant thereof, adequate to produce antibody to protect said individual from infection, particularly HSV-2 infection. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises, through gene therapy, delivering a gene encoding HSV-2, or an antigenic fragment or a variant thereof, for expressing HSV-2, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody to protect said individual from disease.

A further aspect of the invention relates to an immunological composition which, when introduced into a host capable or having induced within it an immunological response, induces an immunological response in such host to HSV-2 or a protein coded

therefrom, wherein the composition comprises a recombinant HSV-2 or protein coded therefrom comprising DNA which codes for and expresses an antigen of said HSV-2 or protein coded therefrom.

5 The HSV-2 or a fragment thereof may be fused with a co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. This fused recombinant protein, preferably further comprises an antigenic co-protein, such as Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilise the protein and facilitate production and purification thereof. Moreover, the co-protein may act  
10 as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

The present invention also includes a vaccine formulation which comprises the immunogenic recombinant protein together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for  
15 example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending  
20 agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known  
25 in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Whilst the invention has been described with reference to ascertain HSV-2 polypeptide, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins (for example, having sequence homologies of 75% or greater)  
30 with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

#### Compositions:

The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or the inhibitors. Thus, the polypeptides of the present invention  
35 may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells,



tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration.

**Kits:**

The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, reflecting approval by the agency of the manufacture, use or sale of the product for human administration.

**Administration:**

Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

The pharmaceutical compositions generally are administered in an amount effective for treatment or prophylaxis of a specific indication or indications. It will be appreciated that optimum dosage will be determined by standard methods for each treatment modality and indication, taking into account the indication, its severity, route of administration, complicating conditions and the like.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

The composition of the invention may be administered by injection to achieve a systemic effect against relevant virus shortly before insertion of an in-dwelling device. Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent viral reactivation.

Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response.

A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks.

With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

In order to facilitate understanding of the following example certain frequently occurring methods and/or terms will be described.

#### Example 1

##### **Preparation of ultra-purified Herpes simplex 2 virus DNA:**

This protocol describes the preparation of herpes simplex virus type 2 strain SB5 DNA for sequencing. It is the combination of two protocols, both of which have been modified. Part one describes the crude isolation of the viral DNA from host cell DNA (Hirt, B., *J. Mol. Biol.* 26: 365-369. (1967)), and part two describes the ultra-purification of the viral DNA through a cesium chloride (CsCl) gradient (Vinograd J, et al., *Proc. Nat'l. Acad. Sci.(USA)* 2:902-910(1963)).

##### **I. Separation of viral DNA from host DNA (modified from Hirt<sup>1</sup>)**

Confluent monolayers of Vero cells (ATCC CCL 81) previously seeded into roller bottles ( $1 \times 10^8$  cells/bottle), were infected with HSV-2 strain SB5 at an MOI = 0.01 in

HBSS. After one hour, the virus inoculum was removed and normal media was added (DMEM, 10% FCS).

Approximately 40-48 hours post- infection, infected monolayers were harvested by scraping, and placed in 10ml of cold 1x PBS. For subsequent steps, three roller bottles of infected cells were combined ( $3 \times 10^8$  cells) The cells were spun at 2000g x 5 minutes. The supernatant was removed and to the cell pellet, 25ml of DNA extraction buffer was added (0.25% Triton X-100, 10mM EDTA, 10mM Tris pH 8.0).

The lysate was mixed at room temperature for 10 minutes. Then to the lysate, 1ml of 5M NaCl (0.2M final concentration) was added and allowed to mix another 15 minutes.

The lysate was centrifuged at 10,000g for 30 minutes at 4°C. The supernatant, which contains the viral DNA, was saved and the pellet, which contains mostly chromosomal DNA, was discarded.

To the supernatant, SDS was added to 0.5% final conc. and Proteinase K to 150ug/ml final conc. This was incubated 2 hours at 45°C.

After two hours, 2.5 volumes of 100% ethanol were added. Viral DNA was precipitated overnight at -20°C.

The precipitate was centrifuged at 10,000g for 30 minutes at 4°C. The pellet was washed once with 70% ethanol and air dried for 30 minutes. Then the pellet was resuspended in 250ul of TE (10mM Tris, pH 7.5, 2mM EDTA).

RNase A was added to a final concentration of 10ug/ml and incubated at 37°C for one hour.

SDS and Proteinase K were then added (as above) and incubated overnight at 37°C.

The DNA was phenol extracted 2x, chloroform extracted 1x, and 1/10 volume 3M sodium acetate and 2.5 volumes of 100% ethanol to precipitate were added and allowed to precipitate overnight at -20°C. The next day, The precipitate was spun down at 15,000g x 20 minutes. The pellet was washed 1x with 70% ethanol, briefly air dried and resuspended in 1ml of TE.

## II. Ultrapurification of the viral DNA through a CsCl gradient (modified from Vinograd, et al. supra)

A cesium chloride solution of 57% w/w with the prepared DNA from above was made as follows:

To the 1ml of viral DNA prepared above, 9ml of TE was added for a total of exactly 10ml. To this, 13.26g of CsCl was added and dissolved. This solution was added to ultracentrifuge tubes and spun in a VTi 40 rotor at 35,000 rpm for 72 hours at 25°C.

After centrifugation, the tube was mounted on a gradient collector and through a hole pierced in the bottom, 15 drop fractions were collected.

The refractive index of every fourth tube was determined on a refractometer. The viral DNA lies between refractive indices = 1.403-1.401. Density range for HSV DNA from Goldin A.L, *et al.*, *J. Virol.* \_\_\_\_: 50-58. Boyant density ( $\rho$ ) =  $a \eta^{25^\circ} - b$ , where coefficients a and b are 10.8601 and 13.4974 respectively for CsCl,  $\eta$  = refractive index. (Isco tables, a handbook of data for biological and physical scientist, Isco, Inc. Lincoln, NE, ninth ed. 1987).

The appropriate fractions were pooled and dialyzed against 3L of TE with frequent changing overnight.

The final DNA prep was concentrated by precipitating with 1/10 volume 3M sodium acetate and 2.5 volumes of 100% ethanol. The DNA was resuspended in TE and the OD<sub>260/280</sub> reading taken.

The DNA was then subjected to sequencing as provided in Sambrook, J et al. (1989) Chapter 13, *supra*; or by automated DNA sequencing as per manufacturer's protocols, e.g., Applied Biosystems/Perkin Elmer, Foster City, CA.

Certain preferred individual polynucleotide and polypeptide sequences of the invention are summarized in the following Tables. Tables 1, 2 and 3 represent three different sequencing efforts. Table 4 represents polypeptides encoded by ORFs from Table 3.

**Table 1** provides polynucleotides of the invention and polypeptides encoded by ORFs, wherein the polynucleotide start and end position for each ORF is indicated by sequence numbers which correlate to the the polynucleotide sequence referred to above each given polypeptide in the Table. Additionally, each ORF-encoded polypeptide sequence is labeled with the Contig number matching the Contig number of the polynucleotide sequence from which it was encoded. For ORF sequences wherein the start polynucleotide number is larger than the end polynucleotide number, translation of that polypeptide initiates on the nucleotide strand which is complementary to the strand depicted in the Table. In many cases there is more than one ORF mapped to an individual Contig. Contig assembly was performed using the publicly-available Phrap program, P.Green, University of Washington, WA., U.S.A. ORF prediction was accomplished using the publicly-available GenMark program, Georgia Tech Research Corp., Georgia Tech, GA, U.S.A. Homologies of the polypeptide sequences to known proteins are also indicated. These homologies were

determined using the public database Mpsrch\_pp, release 2.1 by J.Collins, Biocomputing Research Unit, University of Edinburgh (distributed by IntelliGenetics, Inc.).

**Table 2**, obtained from a separately-performed sequencing, provides polynucleotides of the invention and polypeptides encoded by ORFs, wherein the polynucleotide start and end position for ORFs are indicated by sequence numbers which correlate to the polynucleotide sequence referred to above each given polypeptide in the Table. Each ORF-encoded polypeptide sequence is labeled with a Contig number matching the Contig number of the polynucleotide sequence referred to above it, from which it was encoded. For ORF sequences wherein the nucleotide start number is larger than the end number, translation of that polypeptide initiates on the nucleotide strand which is complementary to the strand depicted in the Table. Contig assembly was accomplished using the publicly-available Sequencer 3.0, Gene Codes Corp., Ann Arbor MI, USA, software program. ORF prediction was done using the publicly-available GenMark program (see Table 1). Homologies of the polypeptide sequences to known proteins are indicated. These homologies were determined using the publicly-available Mpsrch program (see Table 1).

**Table 3** obtained from a separately-performed sequencing, provides polynucleotides of the invention and polypeptides encoded by ORFs, wherein the polynucleotide start and end positions for each ORF is indicated by sequence numbers which correlate to the polynucleotide sequence referred to above that polypeptide in the Table. Each ORF-encoded polypeptide sequence is labeled with a Contig number matching the Contig number of the polynucleotide sequence appearing above it from which it was encoded. For ORF sequences wherein the start polynucleotide number is larger than the end number, translation of that polypeptide initiates on the nucleotide strand which is complementary to the strand depicted in the Table. Contig assembly was performed using the publicly-available Phrap program, (see Table 1). ORF prediction was accomplished using the publicly-available GenMark software program (see Table 1). Homologies of the polypeptide sequences to known proteins are indicated. These homologies were determined by comparison with public database Mpsrch\_pp (see Table 1).

**Table 4** provides ORF sequences of polypeptides encoded by the polynucleotide sequences of Table 3 which were predicted by the GenMark program (see Table 1) as having more than a single start site (N-terminal methionyl residue). The Contig numbers and polynucleotide start and end sites for these ORFs correlate to the Contig numbers and polynucleotide sequence numbers of Table 3.

TABLE 1

- 5 [SEQ ID NO:1] = Contig ID 100
- [SEQ ID NO:2]  
ORF # = 1 from Contig ID 100  
ORF start site = 1808
- 10 ORF end site = 3  
ORF sequence:  
MVLMGRLRNAPESLTVMFCAAIRVAPVTTQSRTSLRVCTHVLFPPDPALPVMRYAANGNSR  
SGRPVGTSKAATSRNHCRRGTCVTSSCCCESSRMRAMIGWTPCMDVKFKNASSLNRTAGL  
APGCCGGGPGARTSREPSPPDAAMAAQRARAPAMRTRGGDAALCAPEDGWVKVHPTPGTM  
15 LFREILLGQMGYTEGQGVYNVRSSEAATRQLQAAIFHALLNATTYRDLEEDWRRHVVAR  
GLQPQRLVRRYRNAREGDIAGVAERVFDTWRCTLRITLLDFAHGVVNCFAPGGPSGPTSF  
PKYIDWLTCGLVPILRKTREREATQRLGAFLRQHTLPRQLATVAGAAERAGPGLLELAV  
AFDSTRMAEYDRVHIYNNHRRGEWLVRDPVSGQRGECLVLCPLWTGDRLVFDSVPVQRLC  
PEIVACHALREHAHICRLRNTASVKVLLGRKSDSERGVAGAARVVKALGEDDETKAGSA  
20 ASCLVRLIINMKGMRHVGDINDTVRAYLDEAGGHLIDTPAVDHTLPGFGKGGTGRGSAAQ  
DPGARPQQLRQAFQTAVVNNINGMLEGYINNLFGTIERLRETNAGLATQLQARGGSSRST  
AX
- 25 Gene matched: gi|136794|sp|P10190|UL06\_HSV11  
Gene name: VIRION PROTEIN UL6. gi|73994
- 30 [SEQ ID NO:3]  
ORF # = 2 from Contig 100  
ORF start site = 1378  
ORF end site = 4023  
ORF sequence:
- 35 MAASGGEGSRDVRAPGPPPPQPGARPAVRFRDEAFLNFTSMHGVQPIIARIRELSQQQLD  
VTQVPRLQWFRDVAALVPTGLPLREFPFAAYLITGNAGSGKSTCVQTLNEVLDCVVTGA  
TRIAAQNMVYKLSGAFLSRPINTIFHEFGFRGNHVQAQLGQHPYTLASSPASLEDLQRRD  
LTYTWEVILDITKRALAAHGGEARNEFHALTAEQTLGLGQGALTRLASVTHGALPAFT  
RSNIIVIDEAGLLGRHLLTTVVYCWMMINALYHTPQYAGRLRPVLVCVGSPTQTASLEST  
40 FEHQKLRCVSRQSENVLTYLICNRTLREYTRLSHSWAIFINNRCVEHEFGNLMKVLEYG  
LPITEEHMQFVDRFVVPESYITNPANLPGWTRLFSSHKEVSAYMAKLHAYLKVTREGEFV  
VFTLPVLTFVSVKEFDKYRRLTQQPTLTMEKWITANASRITNYSQSQDQDAGHVRCEVHS  
KQQLVVARNDITYVLNSQVAVTARLRKMVFGFDGTFRTFEAVLRDSDSVKQTQGETSVEFA  
YRFLSRLMFGGLIHFYNFLQRPGLDATQRTLAYGRLGELTAELLSLRDAAGASATRAAD  
45 TSDRSPGERAFNFKHLGPRDGGPDDFPDDDLVIFAGLDEQQLDVFYCHYALEEPETTAA

VHAQFGLLKRAFLGRYLILRELFGEVFESAPFSTYVDNVIFRGCELLTGSPRGGLMSVAL  
QTDNYTLMGYTTYTRVFATAEELRRRHATAGVAEFLEESPLPYIVLRDQHGFMSSVNTNIS  
EFVESIDSTELAMAINADYGISSKLAMTITRSQGLSLDKVAICFTPGNLRNLNSAYVAMSR  
TTSSEFLHMNLNPLRERHERDDVISEHILSALRDPNVVIVY\*

5

Gene matched: gi|74000|pir||WMBEU5

Gene name: gene UL5 protein - human herpesvirus 1

10

[SEQ ID NO:4]

ORF # = 3 from Contig 100

ORF start site = 4090

ORF end site = 4695

15

ORF sequence:

MGNPQTTIAYSLHHPRASLTALPDAAQVVHVFESESGTRAVLTRGRARQDRLPRGGVVIQH  
TPIGLLVIIDCRAEFCAIRFIRASTQRLERWWDAMHMYAIPFDSWVSSSHGESVRSATAG  
ILTVVWTPDTIYITATIIYGTAPAAARGCDNAPLDVRPTTPPAPVSPTAGFPANTTDLV  
EVLREIQISPTLDDADPTPGT\*

20

Gene matched: gi|136788|sp|P28280|UL04\_HSV2H

Gene name: PROTEIN UL4. gi|73890|pir||W

25

[SEQ ID NO:5]

ORF # = 4 from Contig 100

ORF start site = 5413

ORF end site = 4895

ORF sequence:

30

VGPLDGEPRDAISPLTSSVAGDPPGADGPVYTFDTLFMVSSIDELGRRQLTDTIRKDLR  
LSLAKFSIACKTSSFSGTAAARQKRGAAPPQRTCVPRSNKSLQMFVLCRANAAQVREQL  
RAVIRSRKPRKYTRSSDGRCLCPAVPVFVHEFVSSEPMRLHRDNVMLSTEPD\*

35

Gene matched: gi|330308

Gene name: (L02638) nuclear phosphoprotein [Herpes simplex v

[SEQ ID NO:6]

ORF # = 5 from Contig 100

40

ORF start site = 6656

ORF end site = 5652

ORF sequence:

MKRARSRSPPSPSRPSSPFRTPPHGGSPRREVGAGILASDATSHVCIASHPGSGAGYPTR  
LAAGSAVQRRRPRGCPGVMFSASTTPEQPLGLSGDATPPLPTSVPLDWAAFRAFLIDD  
45 AWRPLLEPELANPLTARLLAEYDRRCQTEEVLPREDVFSWTRYCTPDDVRVVIIGQDPY

HHPGQAHGLAFSVRADVPVPPSLRNVLA AVKNCYPDARMSGRGCLEKWARDGVLLLNNTTL  
TVKRGAAASHSKLGWDRFVGGVVRRLAARRPGLVFM LWGAHAQN A IRPDPRQH YVLKFSH  
PSPLSKVPFGTCQHFLAANRYLETRDIMPIDWSV\*

5

Gene matched: gi|330306

Gene name: (M25410) uracil-DNA glycosylase [Herpes simplex v

10

[SEQ ID NO:7]

ORF # = 6 from Contig 100

ORF start site = 7080

ORF end site = 6529

15

ORF sequence:

VPCMRTPADDVSWRYEAPSVIDYARIDGIFLRYHCPGLDTFLWDRHAQRAYLVNPF LFAG  
GFLEDLSH SVFPADTQETTTRRALYKEIRDALGSRKQAVSHAPVRAGCVNFDYSRTRRCV  
GRRDLRPANTTSTWEPPVSSDDEASSQSKPLATQPPVLALSNA PPRRVSPTRGRRRHTRL  
RRN\*

20

Gene matched: gi|136776|sp|P28278|UL01\_HSV2H

Gene name: GLYCOPROTEIN L PRECURSOR. gi

25

[SEQ ID NO:8] = Contig ID 101

[SEQ ID NO:9]

30

ORF # = 1 from Contig 101

ORF start site = 351

ORF end site = 1259

ORF sequence:

35

MIRRRGNVEIRVYYESVRPSRSRSHLKPSD HQEFPGHHVSPGSPGFPESPGNREFHDLPE  
NPGSRAYPGTRDPHDPHGCPGSLDPHG NPAQPA GLPSPVPYAPLGSPDPSSPRQRTYVLP  
RVGIRNAPASDTRAPKRAHSRHRADRPPE SPGSELYPLNAQALAH LQMLPADHRAFFRTV  
IEVSRLCALNTHDPPPLAGARVGQEAQLVHTQWLRANRESSPLWPWR TAAMNFIAAAAP  
CVQTHRHMHDLLMACAFWCCLAHASTCSYAGLYSAHCQHLFRAFGCGPPVL TTSRGQGGW  
CN\*

40

Gene matched: gi|757866

Gene name: (X02138) 34K (Us10) (aa 1-284) [Human herpesvirus

45

[SEQ ID NO:10]



ORF # = 2 from Contig 101  
ORF start site = 2140  
ORF end site = 1871  
ORF sequence:  
5 MTSRPADQDSVRSSASVPLYPAASPVPAAEAYYSESEDEAANDFLVRMGRQQSVLRRRRRR  
TRCVGLVIACLVVALLSGGFGALLVWLLR\*

Gene matched: gi|135568|sp|P06481|TEGP\_HSV11  
10 Gene name: TEGUMENT PHOSPHOPROTEIN US9

[SEQ ID NO:11]  
ORF # = 3 from Contig 101  
ORF start site = 2377  
15 ORF end site = 2240  
ORF sequence:  
VALHAVDAPSQFVTWLAVRWLRGAVGLGAVLCGIAFYVTSIARGA\*

20 Gene matched: gi|477669|pir||B45696  
Gene name: 23-29K immunoreactive epitope dispens

[SEQ ID NO:12]  
ORF # = 4 from Contig 101  
25 ORF start site = 3572  
ORF end site = 2529  
ORF sequence:  
VAPPRHHRVIEVSHVRGVTVMETPEAIMFAPGETFETKVSIIHAVAHDGPGYAMDVVWM  
RFDVPSSCAEMRIYEACLYHPQLPECLSPADAPCAVSSWAYRLAVRSYAGCSRTTPPPRC  
30 FAEARMEPVPGWLASTVNLEFQHASPQHAGLYLCVVYVDDHIIHAWGHMTISTAAQYRN  
AVVEQHLPPQRQPEPVEPTRPHVRAPPPAPSARGPLRLGAVLGAALLAALGLSAWACMTC  
WRRRSWRAVKSRASATGPTYIRVADSELYADWSSDSEGERDGSWQDPPERPDSPSTNGS  
GFEILSPTAPSVYPHSEGRKSRRLPTTFGSGSPGRRHSQASYSSVLW\*

35 Gene matched: gi|138240|sp|P04488|VGL\_HSV11  
Gene name: GLYCOPROTEIN E PRECURSOR. gi

40 [SEQ ID NO:13]  
ORF # = 5 from Contig 101  
ORF start site = 4176  
ORF end site = 3460  
ORF sequence:  
45 MARGAGLVFFVGWVVSCLAAAPRTSWKRVTSGEDVLLPAPAGPEERTRAHKLWAAEP

LDACGPLRPSWVALWPARRVLETVVDAACMRAPPLAIAYSPPFPAGDEGLYSELAWRDR  
 VAVVNESLVIYGALETDSGLYTLVSVGLSDEARQVASVVLVVEPAPVPTPTPDDYDEEDD  
 AGVSERTPVSVPPPTPPRWSPRGPPEAPSCYPRGVPRARGNGPYGDPGGHYVCPRGDV\*

5

Gene matched: gi|138241|sp|P13289|VGLE\_HSV2  
 Gene name: GLYCOPROTEIN E PRECURSOR. gi|

[SEQ ID NO:14]

10 ORF # = 6 from Contig 101

ORF start site = 5796

ORF end site = 4495

ORF sequence:

VYLWARVGGWLGYLGGTWTPHKGSLEGGKLGQFIGRERGARTAVPTISHRAHSHLDPSDP  
 15 GMPGRSLQGLAILGLWVCATGLVVRGPTVSLVSDSLVDAGAVGPQGFVEEDLRVFGELHF  
 VGAQVPHTNYDGIIELFHYPLGNHCPRVVHVVTLTACPRRPAVAFTLCRSTHHAHSPAY  
 PTLELGLARQPLLRVTATRDYAGLYVLRVWVGSATNASLFLGVALSANGTFVYNGSDY  
 GSCDPAQLPFSAPRLGPSSVYTPGASRPTPPRTTSPSSPRDPTAPAGDTGTPAPASGER  
 APPNSTRSASESRHRLTVAQVIQIAIPASIIAFVFLGSCICFIHRCQRRYRRPRGQIYNP  
 20 GGVSCAVNEAAMARLGAELRSHPNTPPKPRRRSSSTTMPSLTSIAEESFPGPVVLLSVS  
 PRPRSGPTAPQEV\*

Gene matched: gi|138328|sp|P06764|VGLI\_HSV23

25 Gene name: GLYCOPROTEIN I. gi|73722|pir

[SEQ ID NO:15]

30 ORF # = 7 from Contig 101

ORF start site = 7017

ORF end site = 5815

ORF sequence:

VCIAYHGMGRLTSGVGTAAALLVAVGLRVVCAKYALADPSLKMADPNRFRGKNLPVLDQL  
 35 TDPGPKRVYHIQPSLEDPFQPPSIPITVYYAVLERACRSVLLHAPSEAPQIVRGASDEA  
 RKHTYNLTIAWYRMGDNCAIPITVMEYTECPYNKSLGVCPIRTQPRWSYYSFSAVSEDN  
 LGFLMHAPAFETAGTYLRLVKINDWTEITQFILEHRARASCKYALPLRIPPAACLTISKAY  
 QQGVTVDSIGMLPRFIPENQRTVALYSLKIAGWHGPKPPYTSTLLPPELSDTTNATQPEL  
 VPEDPEDSALLEDPAGTVSSQIPPNWHIPSIQDVAPHHAPAAPSNPGLIIGALAGSTLAV  
 40 LVIGGIAFWVRRRAQMAPKRLRLPHIRDDDAPPSHQPLFY\*

Gene matched: gi|419141|pir||E43674

Gene name: US6 protein - human herpesvirus 2 (st

45

- [SEQ ID NO:16]  
ORF # = 8 from Contig 101  
ORF start site = 7553  
5 ORF end site = 7440  
ORF sequence:  
VGGLCLMILGMACLLEVLRRRLGRELARCCPHAGQFAP\*
- 10  
Gene matched: gi|137132|sp|P13293|VGLJ\_HSV2  
Gene name: GLYCOPROTEIN J. gi|419140|pir
- 15 [SEQ ID NO:17] = Contig ID 102
- [SEQ ID NO:18]  
ORF # = 1 from Contig 102  
ORF start site = 1502  
20 ORF end site = 465  
ORF sequence:  
VCPPTPTNMAVVCSGSLRLRPFHPPSPSFFVLRALIRAGPGPFAASPRAPSGPGCGMCRG  
DSPGVAGGSGEHCLGGDDGDDGRPRLACVGAIARGFAHLWLQATTLGFGSVVLSRGPYA  
DAMSGAFVIGSTGLGFLRAPPAFARPPTRVCWLRLVGGGAVALWSLGEAGAPPVPGP  
25 ATQCLALGAAYAALLVLADDVHPLFLLAPRPLFVGTGVLVVVGLTIGGSARYWWIDPRAA  
AALTAADVAGLTGTAAGDSFSKACPRHRRFCVVS AVESPPPRYAPEDAERPTDHGPLLPS  
THHQRSRPRVCGDGAARPENIWVPVVT FAGALALACAARGW WERS\*
- 30 Gene matched: gi|136909|sp|P10227|UL43\_HSV11  
Gene name: MEMBRANE PROTEIN UL43. gi|73
- [SEQ ID NO:19]  
ORF # = 2 from Contig 102  
35 ORF start site = 2996  
ORF end site = 1584  
ORF sequence:  
MAHLEPGGAAAAPLSEDAIPSPRERTEDWPPCQIVLQGAELNGILQAFAPLRTSLDLSLLV  
VGDRGILVHNAIFGEQVFLPLDHSQFSRYRWGGPTAAFLSLVDQKRSLLSVFRANQYPDL  
40 RRVELTVTGQAPFRTLQRIWTTASDGEAVELASETLMKRELTSFAVLLPQGDPDVQLRL  
TKPQLTKVVNAVGDDETAKPTTFELGPNKGFSVFNARTCVTFAAREEGASSSTSAQVQILT  
SALKKAGQAAANAKTVYGENTHRTFSVVVDDCSMRAVLRLQVGGGT LNFFLTADVP SVC  
VTATGPNAVSAVFLLKPQRVCLNWLGRTPGSSTGSLASQDSRAGPTDSQDFSSEP DAGDR  
GAPEEEGLEQARVPPAFPEPPGTKRRHAGAEVVPADDATKRPKTGVPAPTRAESPPLS  
45 ARYGPEAAEGGGDGGRYAWYFRDLQTDASPSPLSAFRGPQRPPYGFGLP\*

Gene matched: gi|136905|sp|P10226|VPAP\_HSV11  
Gene name: POLYMERASE ACCESSORY PROTEIN

5

[SEQ ID NO:20]  
ORF # = 3 from Contig 102  
ORF start site = 3490  
10 ORF end site = 4152  
ORF sequence:  
MGLFGMMKFAQTHHLVKRRGLRAPEGYFTPIAVDLWNVMTLVVKYQRRYPSYDREAITL  
HCLCSMLRVFTQKSLFPFIVTDRGVECTEPVVFGAKAILARTTAQCRTDEEASDVDAASP  
PFPHHRLQAQFPFQHAPPRARLRPGGPGERGPPAQARRPPGARPRSRPCAWLTCSVSAF  
15 CGRWGTPTSTRVSWRPTTPARTSIIPTRSTCIPRIPISC\*

Gene matched: gi|549322|sp|P36699|VHS\_HSV2G  
Gene name: VIRION HOST SHUTOFF PROTEIN.

20

[SEQ ID NO:21]  
ORF # = 4 from Contig 102  
ORF start site = 4122  
25 ORF end site = 4970  
ORF sequence:  
VHTTDTDLLMGCDIVLDISTGYIPTIHCARDLLQYFKMSYPQFLALFVRCHTDLHPNNTY  
ASVEDVLRECHWTAPSRSQARRGARRERANSRSLSEMP TLTAAPVGLETRISWTEILAQQ  
IAGEDDYEDPPLQPPDVAGGPRD GARSSSSEILTPPELVQVPNAQRVAEHRGYVAGRRR  
30 HVIHDAPEALDWLPDMTIAELVEHRYVKYVISLISPKERGPWTLKRLPIYQDLRDEDL  
ARSIVTRHITAPDIADRFLAQLWAHAPPPAFYKDV LAKFWDE\*

Gene matched: gi|549322|sp|P36699|VHS\_HSV2G  
Gene name: VIRION HOST SHUTOFF PROTEIN.

35

[SEQ ID NO:22]  
ORF # = 5 from Contig 102  
ORF start site = 6266  
40 ORF end site = 5253  
ORF sequence:  
MDPAVSPASTDPLDTHASGAGAAPIPVCPTPERYFYTSQCPDINHLRSL SILNRWLETEL  
VFVGDEEDVSKLSEGELGFYRFLFAFLSAADDLV TENLGGLSGLFEQKDILHYYVEQECI  
EVVHSRVYNIIQLVLFHNNDQARRAYVARTINHPAIRVKVDWLEARVRECDSIPEKFILM  
45 ILIEGVFFAASFAAIAYLRTNNLLRVTCQSNDLISRDEAVHTTASCIYNNYLGGHAKPE

AARVYRLFREAVDIEIGFIRSQAPTDSSILSPGALAAIENYVRFSAADRLGLIHMQPLYS  
 APAPDASFPLSLMSTDKHTNFFECRSTSYAGAVVNDL\*

5 Gene matched: gi|132624|sp|P03174|RIR2\_HSV23

Gene name: RIBONUCLEOSIDE-DIPHOSPHATE R

[SEQ ID NO:23]

10 ORF # = 6 from Contig 102

ORF start site = 9861

ORF end site = 6319

ORF sequence:

VIRRPVRPFGRTAHPASHGPAAVSVHRVRATVTLVPMANRPAASALAGARSPSERQEPRE  
 15 PEVAPPGGDHVFCRKVSGVMVLSSDPPGPAAYRISDSSFVQCGSNCSMIIDGDVARGHLR  
 DLEGATSTGAFVAISNVAAGDGRTAVVALGGTSGPSATTSVGTQTSGEFLHGNPRTPEP  
 QGPQAVPPPPPPFPWGHECCARRDARGGAEKDVGAESWSDGPSSDSETEDSDSSEDT  
 GSGSETLSRSSSIWAAGATDDDDSDSDSRSDSVQPDVVRRRWSGDPAPVAFPKPRRPG  
 DSPGNPGLGAGTGPGSATDPRASADSDSAHAAAPQAEVAPVLDSQPTVGTDPGYPVPLE  
 20 LTPENAEAVARFLGDAVDREPALMLEYFCRCAREESKRVPRTFGSAPRLTEDDFGLLNY  
 ALAEMRRICLDLPPVPPNAYTPYHLREYATRLVNGFKPLVRRSARLYRIILGILVHLRIRT  
 REASFEEWMRSKEVDLDFGLTERLREHEAQLMILAQALNPYDCLIHSTPNTLVERGLQSA  
 LKYEEFYLRFGGHYMESVFQMYTRIAGFLACRATRGMRHIALGRQGSWWEMFKFFFHRL  
 YDHQIVPSTPAMNLGTRNYYTSSCYLVNPQATTNQTALRAITGNVSAILARNGGIGLCM  
 25 QAFNDASPGTASIMPALKVLDSLVAAHNKQSTRPTGACVYLEPWHSDVRAVLRMKGVLAG  
 EEAQRCDNIFSAWMPDLFFKRLIRHLDGEENVTWLSLFRDRTSM SLADFHGEFEKLYEH  
 LEAMGGETIPIQDLAYAIVRSAAATTGSPFIMFKDAVNRHYIYNTQGAAIAGSNLCTEIV  
 HPSSKRSSGVCNLGSVNLARCVSRRTDFGMLRDAVQACVLMVNIMIDSTLQPTPQCARG  
 HDNLRSMGIGMQGLHTACLKMGDLLESAEFRDLNTHIAEVMLLAAMKTSNALCVRGARPF  
 30 SHFKRSMYRAGRFHWERFSNASPRYEGEWEMLRQSMKHLRNSQFIALMPTAASAQISD  
 VSEGFAPLFTNLF SKVTRDGETLRPNTLLKELERTFGGKRLLDAMDGLEAKQWSVAQAL  
 PCLDPAHPLRRFKTAFDYDQELLIDL CADRAPYVDHSQSMTLYVTEKADGTL PASTLVRL  
 LVHAYKRGLKTGMYCKVRKATNSGVFAGDDNIVCTSCAL\*

35

Gene matched: gi|330199

Gene name: (M12700) ribonucleotide reductase large subunit (

40

[SEQ ID NO:24]

ORF # = 7 from Contig 102

ORF start site = 11144

ORF end site = 10323

45 ORF sequence:

5 VRRRLRCARRRRGGPGPHHDQLRRDAGRGAAGPVFRMPARHGPHARVSPRGHAVFRGASV  
VVTQDELASVTAVCSGPQEATHTGHPGRPCSAVTIPACAFVDLDAELCLGGPGAFLYLV  
FTYRQCRDQELCCVYVVKSQLPPRGLEAALERLFGRLRITNTIHGAEDMTPLPPNRNVDF  
PLAVLAASSQSPRCSASQVTNPQFVDRLYRWQPDLRGRPTARTCTYAAFAELGVMPDNSP  
RCLHRTERFGAVGVPVVILEGVVWRPGGWRACA\*

Gene matched: gi|139176|sp|P22486|VP19\_HSV2G

Gene name: CAPSID ASSEMBLY AND DNA MATU

10

[SEQ ID NO:25]

ORF # = 8 from Contig 102

ORF start site = 11722

ORF end site = 10667

15

ORF sequence:

MKTKPLPTAPMAWAESAVETTTSPRELAGHAPLRRVLRPPIARRDGPVLLGDRAPRRTAS  
TMWLLGIDPAESSPGTRATRDDTEQAVDKILRGARRAGGLTVPGAPRYHLTRQVTLTDLCL  
QPNARAGALLLALRHPTDLPHLARHRAPPGRQTERLAEAWGQLEASALGSGRAESGCA  
RAGLVSFNFLVAACAAAYDARDAEAVRAHITTNYGGRAGARLDRFSECLRAMVHTHV  
20 PHEVMRFFGGLVSWSHRTSWLASPPSAADPRRPHTPATRAGFVRPLPSRPAPLWTWTPSC  
AWGALGRRSCTWFSPTDSAGTRSSVACTWSRASPRADWRRPSSGCSGASG\*

Gene matched: gi|139176|sp|P22486|VP19\_HSV2G

25

Gene name: CAPSID ASSEMBLY AND DNA MATU

[SEQ ID NO:26] = Contig ID 103

30

[SEQ ID NO:27]

ORF # = 1 from Contig 102

ORF start site = 3308

ORF end site = 693

ORF sequence:

35

MAETMNVATCTHQTHHAARAPGATSAPGAASGDPLGARRPIGDDECEQYTSSVSLARMLY  
GGDLAEWVPRVHPKTTIERQQHGPVTFPDASAPTARCVTVVRAPMGSGKTTALIRWLGEA  
IHSPDTSVLVVSCRRSFTQTLATRFAESGLPDFVTYFSSTNYIMNDRPFHRLIVQVESLH  
RVGPNLLNNYDVLVLDEVMSTLGQLYSPTMQQLGRVDALMLRLLRTCPRIIAMDATANAQ  
LVDFLC SLRGEKNVHVIGEYAMPGFSARRCLFLPRLGPEVLQAALRPPGPAGGAPPPDA  
40 PPDAFFGEVEARLAGGDNVCIFLSTVSFAEVVARFCRQFTDRVLLHSLTPPGEVTTWG  
RYRVVIYTTVVTVGLSFDPPHFDMSFAYVKPMNYGPDMSVYQSLGRVRTLKRGELLIYM  
DGSGARSEPVFTPMLLNHVVSASGQWPAQFSQVTNLLCRRFKGRCDASHADAAQARGSRI  
YSKFRYKHYFERCTLACLADSLNHLMLLTLCNMHVRFWGHDAALTPRNFCLFLRGIHFD  
ALRAQRDLRELRCQDPDTSLSAQAAETEEVGLFVEKYLRPDVAPAEVVALMRGLNSLVGR  
45 TRFIYLVLLLEACLRVPMAAHSSAIFRRLYDHYATGVIPTINAAGELELVALHPTLNVA

WELFRLCSTMAACLQWDSMAGGSGRTFSPEDVLELLNPHYDRYMLVFELGHCNVTDGPL  
 LSEDAVKRVADALSGCPPRGSVSETEHALSLFKIIWGELFGVQLAKSTQTFPGAGRVKNL  
 TKRAIVELLDAHRIDHSACRTHRQLYALLMAHKREFAGARFKLRAPAWGRCLRTHASGAQ  
 PNTDIILEAALSELPTAWPMMQGA VNFSTL\*

5

Gene matched: gi|136806|sp|P10193|UL09\_HSV11  
 Gene name: ORIGIN OF REPLICATION BINDIN

10 [SEQ ID NO:28]

ORF # = 2 from Contig 103

ORF start site = 3160

ORF end site = 4590

ORF sequence:

15 VYCSHSSSPMGRAPRGSPPEAAPGADVAPGARA AWWVCVQVATFIVSAICVVGLLVLAS  
 VFRDRFPCLYAPATSYAEANATVEVRGGVAVPLRLDTQSLLATYAITSTLLAAAVYAAV  
 GAVTSRYERALDAARRLAAARMAMPHATLIAGNVCAWLLQITVLLLAHRISQLAHLIYVL  
 HFACLVYLAHFCTRGVLSGTYLRQVHGLIDPAPTHHRIVGPVRAVMTNALLGTLCTA  
 AAASLNTIAALNFNFSAPSMILCLTTLFALLVVSLLLVEGVLCYVVRVLVGPGLGAIA  
 20 ATGIVGLACEHYHTGGYYVVEQQWPGAQTGVRVALALVA AFALAMAVLRCTRAYLYHRRH  
 HTKFFVRMRDTRHRAHSALRRVRSSMRGSRGGPPGDPGYAETPYASVSHAEIDRYGDS  
 DGDPIYDEVAPDHEAELYARVQRPGVPDAEPIYDTVEGYAPRSAGEPVYSTVRRW\*

25 Gene matched: gi|136810|sp|P04288|VIMP\_HSV11  
 Gene name: PROBABLE INTEGRAL MEMBRANE P

[SEQ ID NO:29]

30 ORF # = 3 from Contig 103

ORF start site = 6853

ORF end site = 4784

ORF sequence:

MAAAATPGAKRPADPARDPDSPPKRPRPNSLDLATVFGPRPAPPRPTSPGAPGSHWPQSP  
 35 PRGQPDGGAPGEKARPASPALSEASSGPPTPDIPSPGGAHAIDPDCSPGPPDPMPWSA  
 SAIPNALPPHILAE TFERHLRGLLRGVRSP LAIGPLWARLDYLC SLVVSLEAAGMVDRGL  
 GRHLWRLTRRAPP SAAEAVAPRPLMGFYEAATQ NQADCQLWALLRRGLTTASTLRWGAQG  
 PCFSSQWLTHNASLR LDAQSSAVMFGRVNEPTARNLLFRYCVGRADAGVNDDADAGR FVF  
 HQPGDLAEENVHACGV LMDGHTGMVGASLDILVCPRDPHGYLAPAPQTPLAFYEVKCRAK  
 40 YAFDPADFGAPAASAYEDLMARRSPEAFRAFIR SIPNGVRYFAPGRVPGPEEALVTQDR  
 DWLDSRAAGEKRRCSAPDRALVELNSGVVSEVLLFGVPDLERRTISPVAWSSGELVRRREP  
 IFANRPHPNFKQILVQGNVPRQPLSRLPPATAPGDVPRQAPRGRGGGRDVPPGGRPRSAR  
 RAWRGPRTRGIDPPGGRSDRPDHHPRPRRAGDIPGHPAKQPPGLRRYARQVMGLAFSG  
 ARPCCCRHNVIITDGGEVVSLTAHEFDVVDIESEEEGNFYVPPDMRVVTRAPGPQYRRAS  
 45 DPPSRHTRRRDPDVARPPATLTPPLSDSE\*

Gene matched: gi|119694|sp|P06489|EXON\_HSV2

Gene name: ALKALINE EXONUCLEASE. gi|3302

5

[SEQ ID NO:30]

ORF # = 4 from Contig 103

ORF start site = 5313

10 ORF end site = 4990

ORF sequence:

VTFLGRHRAGAEEGVTFRLEDGRGAPAGRGGAPGPAKASILPDQAVPIALIITPVRVEPG  
IYRDIRRNSRLAFDDTLAKLWASRSPGRGPAAADTTSSSPTAGRSSR\*

15

Gene matched: gi|330252

Gene name: (M11854) 1.9kD ORF [Herpes simplex virus type 2]

20

[SEQ ID NO:31]

ORF # = 5 from Contig 103

ORF start site = 8477

ORF end site = 6894

ORF sequence:

25 VGGRRPGGRMDESGRQRPASHVAADISPQGAHRRSFKAWLASYIHSLSRRASGRPSGPSP  
RDGAVSGARPGSRRRSSFRERLRAGLSRWRVSRSSRRSSPEAPGPAAKLRRPPLRRSET  
AMTSPSPSPSHILSLARIHKLCIPVFAVNPALRYTTLEIPGARSFGGSGGYGEVQLICEH  
KLAVKTIREKEWFAVELVATLLVGEC AFCGGRTHDIRGFITPLGFSLQQRQIVFPAYDMD  
LGKYIGQLASLRATTPSVATALHHCFTDLARAVVFLNTRCGISHLDIKCANVLVMLRSDA  
30 VSLRAVLADFSLVTLNSNSTISRGQFCLQEPDLES PRGFGMPAALT TANFHTLVGHGYN  
QPPELLVKYLNNERAEFNNRPLKHDVGLAVDLYALGQTLLELLVSVVAPSLGVPVTRVP  
GYQYFNNQLSPDFAVALLAYRCVLHPALFVNSAETNTHGLAYDVPEGIRRHRLRNP KIRRA  
FTEQCIN YQRT HKAVLSSVSLPPELRPLLVLVSRLCHANPAARHSLS\*

35

Gene matched: gi|125628|sp|P04290|KR2\_HSV11

Gene name: PROBABLE SERINE/THREONINE-PRO

40

[SEQ ID NO:32]

ORF # = 6 from Contig 103

ORF start site = 8113

ORF end site = 8352

ORF sequence:

45 MAVSDLRRGGRLSLAAGPGASGDERRRDERLTRHRDSPARSRSRKLDRRRDPGRAPETAP



SRGEGPLGRPDARRLRECM\*

Gene matched: gi|93505|pir||B34768

5 Gene name: ORF5 protein - Orf virus (strain NZ2)

[SEQ ID NO:33]

ORF # = 7 from Contig 103

10 ORF start site = 8863

ORF end site = 8204

ORF sequence:

MSRDASHAALRRRLAETHLRAEVYRDQTLQLHREGVSTQDPRFVGAFMAAKAAHLELEAR  
LKSRRARLEMMRQRATCVKIRVEEQAARRDFLTAHRRYLDPALSERLDAADDRDLADQEEQL  
15 EEAANASLWGDGDLADGWMSPGDSDLLVMWQLTSAPKVHTDAPSRPGSRPTYTPSAAGR  
PDAQAAPPETAPSPEPAPGPAADPASGSGFARDCPDGE\*

Gene matched: gi|136823|sp|P04291|UL14\_HSV11

20 Gene name: HYPOTHETICAL UL14 PROTEIN. g

[SEQ ID NO:34]

ORF # = 8 from Contig 103

ORF start site = 8749

25 ORF end site = 10242

ORF sequence:

VYSRPPGVAAGSGPCTPRPGGASRPNVGAGPRGWRLGSSRRPRARPTSDSFAPTPLTSAA  
PASPAMFGQQLASDVQQYLERLEKQRQQKVGVDASAGLTGGDALRVPFLDFATATPKR  
HQTVPVPGVGTLDHCCHEHSPLFSAVARLLFNLSLPAQLRGRDFGGDHTAKLEFLAPELVR  
30 AVARLRFRECAPEDAVPQRNAYYSVLNTFQALHRSEAFRQLVHFVRDFAQLLKTSFRASS  
LAENTGPPKKRAKVDVATHGQTYGTLELFQKMILMHATYFLAAVLLGDHAEQVNTFLRLV  
FEIPLFSDTAVRHFRQRATVFLVPRRHGKTWFLVPLIALSLASFRGIKIGYTAHIRKATE  
PVFDEIDACLRGWFGSSSRVDHVKGETISFSFPDGSRTIVFASSHNTNVSTPSSRGACFP  
GAALPEIDRQTNTARRECGTTRPQPPPPWRGEALLFICNRTMRLWPRPARPRGSSLQTGG  
35 WYTMTERRGATRRWSGG\*

Gene matched: gi|74013|pir||WMBE31

40 Gene name: 38K protein - human herpesvirus 1 gi|5

[SEQ ID NO:35]

ORF # = 9 from Contig 103

ORF start site = 11332

45 ORF end site = 10115

ORF sequence:

VFLFHRSPTPPPKSYTRWPLCFWCVSGPFPTTNMAQRAVWRPQGTGPGPGAAAPPGHARGA  
 PPDARAPDPGPEADLVARIANSVFVWRVVRGDERLKIIFRCLTVLTEPLCQVALPDPDPER  
 ALFCEIFLYLTRPKALRLPSNTFFAIFFFNRERRYCATVHLRSVTHPRTPLLCTLAFGHL  
 5 EAASPPEETPDPAEQLADEPVAHELDGAYLVPTPEPPNPGACCALGPGAWWHLPGGRIY  
 CWAMDDDLGSLCPPGSRARHLGWLLSRITDPPGGGGACAPTAHIDSANALWRAPAVAEAC  
 PCVAPCMWSNMAQRTLAVRGDASLCQLLFQHPVDAVILRQATRRPRITAHLEVVVGRDG  
 AESVIRPTSAGWRLCVLSSYTSRLFATSCPAVARAVARASSDYK\*

10

Gene matched: gi|136829|sp|P10200|UL16\_HSV11

Gene name: PROTEIN UL16. gi|73879|pir||

15 [SEQ ID NO:36]

ORF # = 10 from Contig 103

ORF start site = 12706

ORF end site = 11336

ORF sequence:

20 FLTGYFRVHGIDKLDQRAVQDVTRRHFPVRARPKHAASGVXSGLRQGALVHXAVSGGALGA  
 SDAEAVLAGLEPPGGGRFATPGGPRAAGDDVLNDVLTLPVGTAKPRSLVEWLDRGWEPLA  
 GGDRPDWLWSRRSISVLRHHYGTQRFVVVSYKNSVAWGGRTRRPPLLSSYLATALTEA  
 CAAERVVRPHQLSPAAQTALLRRFPALLEGPLRHPRFVLQPFDIAAEVAFVARIQIACLRA  
 LGHSIRAAALQGGPRIFQRLRYDFGPHQSEWLGEVTRRFVLLLENLMRALEGTAPDAFFHT  
 25 AYALAVLAHLGGQGGRRRRRLVPLSDDIPARFADSDAHYAFDYYSTSGDTLRLTNRPIA  
 VVIDGDVNGREQSKCRFMEGSPSTAPHRVCEQYLPGESYAYLCLGFNRRRLCGLVVFPGGF  
 AFTINTAAYLSLADPVARAVGLRFCRGAGTGPGGLVR\*

30 Gene matched: gi|136835|sp|P10201|UL17\_HSV11

Gene name: PROTEIN UL17. gi|73875|pir||

35 [SEQ ID NO:37] = Contig ID 104

[SEQ ID NO:38]

ORF # = 1 from Contig 104

ORF start site = 3027

40 ORF end site = 262

ORF sequence:

VSGRAGDPAGLPAPRGGPTWPMPSGGPPPEVKAGLRADMWGMQYREAXEHQTPDTETV  
 VAGMHPALVVVLKTMFXDAPETPVLVQFFSDHAPTIKAVSNAINAGSAAVATASPAATV  
 DAAVRAHGALADAVSALGAAARDPASPLSFLAALADSAAGYVKATRLALEARGAIDKLTT  
 45 LGSAAADLVFHARRACAQPEGDHAALIDAAARATTAARESLAGHEAGFGGLLHAEGTAGD

HSPSGRALQELGKVGATRRRAEELEAAVADLTGKMAAQRARGSSERWAAGVEAALDRVE  
 NRAEFDVVELRRLQALAGTHGYNPRDFRKRAEQALAAAEAVTLALDTAFANPYTPENQ  
 RHPMLPPLAAIHRLGWSAAFHAAAETYADMFRVDAEPLARLLRIAEGLLEMAQAGDGFID  
 YHEAVGRLADDMTSVPGLRRYVPPFFQHGYADYVELRDRLDAIRADVHRALGGVPLDLAAA  
 5 AEQISAARNDEATAELVRTGVTLPCPSEDALVACAAALERVQSPVKNTAYAEYVAFVT  
 RQDTAETKDAVVRAKQQRAEATERVMAGLREALAARERRAQIEAEGLANLKTMLKVAVP  
 ATVAKTLDQARSVAEIAEQVEVLLDQTEKTRELDVPAVIWLEHAQRTFETHPLSAARGDG  
 PGPLARHAGRLGALFDTRRRVDALRRSLEEAEAEWDEVWGRFGRVRGGAWKSPEGFRAMH  
 EQLRALQDTTNTVSGLRAQPAYERLSARYQGVLGAKGAERAEAVEELGARVTKHTALCAR  
 10 LRDEVVRVPWEMNFDALGRLLAEFDAAAADLAPWAVEEFRGARELIQYRMGLYSAYARA  
 GKGALFLFFFFPPLSSFLPHFHFFIHHHHSFTKFFTSSSLHSYHLFPSSIYSIPSISPL  
 YPHSSLSFPSSQFLHIFLSLP\*

15 Gene matched: gi|135576|sp|P10220|TEGU\_HSV11  
 Gene name: LARGE TEGUMENT PROTEIN (VIRI

[SEQ ID NO:39]

ORF # = 2 from Contig 104

20 ORF start site = 3914

ORF end site = 2901

ORF sequence:

VMPVAPPPRGAGGRAPCPALGPEAIHARLEDVRIQARRAIESAIKEYFHRGAVYSAKAL  
 QASDSHDCRFHVASAAVPMVQLLESPLAFDQHTRDVAQRAALPPPPPLATSPQAILLRD  
 25 LLQRGQTLDAPEDLAAWLSVLTDAATQGLIERKPLEELARSINGINDQQARRSSGLAELQ  
 RFDALDAALAQQLDSDAAFVPATGPAPYVDGGGLSPEATRMAEDALRQARAMEAAKMTAE  
 LAPEARSLRERAHAEAMLNDARERAKVAHDAREKFLHKLQGVLRPLPDFVGLKACPAV  
 LATLRASLPRGVDRPGRCRPGAPPRKSRRGCGRTCGG\*

30

Gene matched: gi|221757

Gene name: (D10879) virion protein [Herpes simplex virus typ

35 [SEQ ID NO:40]

ORF # = 3 from Contig 104

ORF start site = 6099

ORF end site = 3643

ORF sequence:

VVTGVRNQFATDLEPGGSVSCMRSSLSFLSLLFDVGPRDVLSAEAEGLVEGGEWTRAA  
 AGSGPPRMCSSIIELPNFLEYPAARGGLRCVFSRVYGEVGFGEPTAGLLETQCPAHTFFA  
 GPWAMRPLSYTLTIGPLGMGLYRDGDTAYLFDPHGLPAGTPAFIAKVRAGDVYPYLTY  
 AHDRPKVRWAGAMVFFVPSGPGAVAPADLTAAALHLYGASETYLQDEFFVERRVAITHPL  
 RGEIGGLGALFVGVPVPRGDGEGSGPVVPALPAPTHVQTPRADRPPEAPRGASGPPNTPQA  
 45 GHPNRPDDVWAAALEGTPPAKPSAPDAAASGPPHAAPPPQTPAGDAAEEAEDLRVLEVG

AVPVGCHRRARYSTGLPKRRRPTWTFPSSVEDLTSGERPAPKAPPKAKKKKSAPKKKAPVA  
 AEPASSPTPIAATVPPAPDTPPQSGQGGGDDGPASPSSPSVLETLGARRPPEPPGADLA  
 QLFEVHPNVAATAVRLAARDAALAREVAACSQLTINALRSPYPAHPGLELCVIFFFERV  
 LAFLIENGARTHTQAGVAGPAAALLDFTLRMPPRKTAVGDFLASTRMSLADVAHRPLIQ  
 5 HVLDKNSQIGRLALAKLVLVARDFIRETDAFYGDLADLDLQLRAAPPANLYARLGKWLE  
 RSRAPHNTLFAPATPTHPEPLLHRIQALAHFARGKKMRVEAEAREMREALYALARGVYSV  
 SQRAGPPDRDARCPPPPGRRRQGPVPARPGPRGHPCAAGGRADPGPPGDRKRDQGVLP  
 SRIQREGPAGQRQPRLSVSRGLGRGRAHGPVAGIATGL\*

10

Gene matched: gi|221757

Gene name: (D10879) virion protein [Herpes simplex virus typ

[SEQ ID NO:41]

15 ORF # = 4 from Contig 104

ORF start site = 6751

ORF end site = 6269

ORF sequence:

MNAHFANEVQYDLTRDPSSPASLIHVIISSECLAAAGVPLSALVRGRPDGGAAANFRVET

20 QTRPHAPGDCTPWRSFAAYVPADAVGAILAPVIPAHPDLLPRVPSAGGLFVSLPVACDA

QGVYDPYTVAALRLAWGPWATCARVLLFSYDELTRYRVCG\*

Gene matched: gi|136835|sp|P10201|UL17\_HSV11

25 Gene name: PROTEIN UL17. gi|73875|pir||

[SEQ ID NO:42]

30 ORF # = 5 from Contig 104

ORF start site = 6781

ORF end site = 8052

ORF sequence:

VPEGAWVGGACARPRGPRAHVRLYAVCFVCPQGIRGQDFNLLFVDEANFIRPDAVQTIMG

35 FLNQANCKIIFVSSTNTGKASTSFLYNLRGADELNNVVTYICDDHMPRVVTHTNATACS

CYILNKPVFITMDGAVVRTADLFLPDSFMQEIIGGQARETGDDRVPVLTKSAGERFLLYRP

STTTNSGLMAPELYVYVDPAFATANTRASGTGIAVVGRYRDDFIIFALEHFFLRALTGSAP

ADIARCVVHSLAQVLALHPGAFRSVRVAVEGNSSQDSAVAIATHVHTEMHRILASAGANG

PGPELLFYHCEPPGGAVLYPFFLLNKQKTPAFEFYIKKFNSGGVMASQELVSVTVRLQTD

40 PVEYLSEQLNNLIETVSPNTDVRMYSGKRNGAADDLMVAVIMAIYLAAPTGIPPAFFPIT

RTS\*

Gene matched: gi|139646|sp|P04295|VTER\_HSV11

45 Gene name: PROBABLE DNA PACKAGING PROTE

[SEQ ID NO:43]  
ORF # = 6 from Contig 104  
5 ORF start site = 9483  
ORF end site = 8392  
ORF sequence:  
VLLSPAPPPLPHGRCPPSLFHHRPGCVALS GPPAPPRSGVSRPGAMITDCFEADIAIPSG  
ISRPDAAALQRCGRVVFLLPTIRRLALADVAHESFVSGGVSPDTLGLLLAYRRRFP  
10 TRVLPTRIVACPVDLGLTHAGTVNLRNTSPVDLCNGDPVSLVPPVFEGQATDVRLES  
DLTLRFPVPLPTPLAREIVARLVARGIRDLNPDPRTPGELPDNLVLYNGARLSLVADVQQL  
ASVNTELRSLVLNMVYSITEGTTLILTLIPRLALSAQDGYVNALLQMQSVTREAAQLIH  
PEAPMLMQDGERRLPLYEALVAWLAHAGQLGDILALAPAVRVCTFDGAADVQSGD  
MAPVIRYP\*  
15  
  
Gene matched: gi|139191|sp|P10202|VP23\_HSV11  
Gene name: CAPSID PROTEIN VP23. gi|7387  
  
20  
[SEQ ID NO:44]  
ORF # = 7 from Contig 104  
ORF start site = 13917  
ORF end site = 9727  
25 ORF sequence:  
VWEGGLPELGLMEPANPPRNPMAPARDPPGYRYAAAMVPTGSILSTIEVASHRR  
LFDFFARVRSDENSLYDVEFDALLGSYCNLTSLVRFLGLSVACVCTKFP  
ELAYMNEGRVQFVEHQPLIARDGPHFVEQPVHNYMTKVIDRRALNAAFSLATEA  
IALLTGEALDGTGISLHRQLRAIQQLARNVQAVLGAFERGTA  
DQMLHVLLEKAPPLALLLPMQRYLDNGRLATRVARAT  
30 LVAELKRSFCDTSFFLGKAGHRRERAEAWLVDLTATQPSVAVPRLTHADTRGR  
PVDGVLVTAAIKQRLQLQSFLKVEDTEADVPTYGEMVLNGANLVTALVMGKAV  
RSLDDVGRHLLMQEEQLEANRETLELESAPQTTRVRADLVAIGDRLVFLEALEK  
RIYAATNPYPYPLVGAMDLTFLVPLGLFNPAMERFAAHAGDLVPAPGHPEPRA  
FPFRQLFFWKGDKHQLRLSMENAVGTVCHPSLMNIDAAVGGVNHD  
PVEAANPYGAYVAAPAGPGADMQRFLNAWRQRLAHGRV  
35 RWVAECQMTAEQFMQPDNANLALHLPADFDFAGVADVELPGGEVPPAGPGAI  
QATWRVVENGNLPLALCPVAFRDARGLELGVGRHAMAPATIAAVRGAFEDRSYPA  
VFYLLQAAIHGSEHVFCALARLVTCITSYWNTRCAAFVNDYSLVSYIVTYLGGDL  
PEECMAVYRDLVAHVEALAQLVDDFTLPGPELGGQAELNHLMRDPALLPPLVWDC  
DGLMRHAALDRHRDCRIDA  
40 GGHEPVYAAACNVATADFNRNDGRLLHNTQARAADAADDRPHRPADWT  
VHHKIYYVLPVAFSRGRCTAGVRFDRVYATLQNMVVEPIAPGEECP  
SDPVTDPAHPLHPANLVANTVNM  
FHNGRVVDGPAMLTQVLAHNMAERTTALLCSAAPDAGANTASTANMRIFD  
GALHAGVLLMAPQHLDHTIQNGEYFYVLPVHALFAGADH  
VANAPNFPALRDLARHVPVPPALGANYFSSIRQPVVQHA  
RESAAGENALTYALMAGYFKMSPVALYHQLKTGLHPGFGFTV  
VRQDRFVTENVLFSERASEAYFLGQLQVARHETGGGV  
SFTLTQPRGNVDLGVGYTAVAATATVRNP  
45 VTDGMGNLPQNFYLGRGAPLLNNAAVYLRNAV  
VAGNRLGPAQLPVFGCAQVPRRAGMD

HGQDAVCEFIATPVATDINYFRRPCNPRGRAAGGVYAGDKEGDVIALMYDHGQSDPARPF  
AATANPWASQRFSGYDGLLYNGAYHLNGASPVLSPCFKFFTAADITAKHRCLERLIVETGS  
AVSTATAASDVQFKRPPGCRELVEDPCGLFQEAYPITCASDPALLRSARDGEAHARETHF  
TQYLIYDASPLKGLSL\*

5

Gene matched: gi|137571|sp|P06491|VCAP\_HSV11  
Gene name: MAJOR CAPSID PROTEIN (MCP) (

10

[SEQ ID NO:45]

ORF # = 8 from Contig 104

ORF start site = 14832

ORF end site = 14164

15

ORF sequence:

MTMRDDVPLLDRELVEAACGGEDGELPLDEQFSLSSYGTSDFFVSSAYSRLPHTQPVF  
SKRVVMFAWSFLVLKPLELVAAGMYGWTGRAVAPACIIAAVLAYYVTLARALLLYVNI  
KRDRLPLSPPVFWGLCVIMGGAALCALVAAAHETFSPDGLFWITASQLLPRTDPLRARS  
LGIACAAGAAMWVAADCFAAFTNFFLARFWTRAILKAPVAF\*

20

Gene matched: gi|136841|sp|P10204|UL20\_HSV11  
Gene name: MEMBRANE PROTEIN UL20. gi|73

25

[SEQ ID NO:46]

ORF # = 9 from Contig 104

ORF start site = 15168

ORF end site = 17081

30

ORF sequence:

VGRQGERWVGGGNEKNTQRATSGMRPELSLKGRPCVTEAVVCPSTDAAIHSGGSSSVRPQ  
PYARAARARATHGSRSRHRQPLLPSSSHHTIPPPSPPRGSPAMELTATTLHHRDVV  
FYVTADRNRAYFVCGGSVYSVGRPRDSQPGEIAKFGLVVRGTGPKDRMVANYVRSELQR  
GLREVRPVGEDEVFLDSVCLLNPNVSSERDVINTNDVEVLDECLAEYCTSLQTS PGVLVT  
35 GVRVRARDRVIELFEHPAIVNISSRFAYTPSPYVFALAQHLRPLSSLEPLVSGLFDGI  
PAPRQPLDARDRTDVITGTRAPRPMAGTGAGGAGAKRATVSEFVQVKHIDRVVSPSVS  
SAPPPSAPDASLPPPGQLQEAPPGPPLRELWWVFYAGDRALEEPHAESGLTREEVRAVHG  
FREQAWKLFSGVAPRAFLGAALALSPTQKLAVYYYLIHRERRMSPFPALVRLVGRYIQR  
HGLYVPAPDEPTLADAMNGLFRDALAAGTVAEQLLMFDLLPPKDVPGSDARADSAALLR  
40 FVDSQRLTPGGSVSPEHVMYLGAFLGVLYAGHGRLAAATHARTLTGVTSLVLTVGDDVDRM  
SAFDRGPAGAAGRTRTAGYLDALLTVCLARAQHGQSV\*

Gene matched: gi|136845|sp|P10205|UL21\_HSV11

45

Gene name: PROTEIN UL21. gi|73866|pir||

[SEQ ID NO:47]  
 ORF # = 10 from Contig 104  
 ORF start site = 19116  
 5 ORF end site = 17302  
 ORF sequence:  
 VYLSPSALKWPVGWTTGGLAFGCDALVRARYGKGFMGVVISMRDSPPAEIIIVPADKT  
 LARVGNPTDENAPAVLPGPPAGPRYRVFVLGAPTPADNGSALDALRRVAGYPEESTNYAQ  
 YMSRAYAEFLGEDPGSGTDARPSLFWRLAGLLASSGFVNAHAHDAIRLSDLLGFLAH  
 10 SRVLAGLAARGAAGCAADSVFLNVSVDPAARLRLEARLGHVAAILEREQSLAAHALGY  
 QLAFVLDSPAAYGAVAPSAARLIDALYAEFLGGRALTAPMVRRALFYATAVLRAPFLAGA  
 PSAEQRERARRGLLITTALCTSDVAAATHADLRAALARTDHQKNLFWLPDHFSPCAASLR  
 FDLAEGGFILDALAMATRS DIPADVMAQQTRGVASVLRWAHYNALIRAFVPEATHQCSG  
 PSHNAEPRILVPITHNASYVVTHTPLPRGIGYKLTGVDVRRPLFITYLTATCEGHAREIE  
 15 PKRLVRTENRRDLGLVGAVFLRYTPAGEVMSVLLVDTDATQQQLAQGPVAGTPNVFSSDV  
 PSVALLLFPNGTVIHLAFTLPIATIAPGFLAASALGVVMITAALAGILRVVRTCVPFPL  
 WRRE\*

20 Gene matched: gi|138316|sp|P08356|VGLH\_HSV1E  
 Gene name: GLYCOPROTEIN H PRECURSOR. gi

[SEQ ID NO:48]  
 25 ORF # = 11 from Contig 104  
 ORF start site = 20070  
 ORF end site = 19117  
 ORF sequence:  
 VVISAGVRGQGWHRISTPPKNGAGRSVLVFGVLPLCFYPHPTPSFGPRLRQQRASDSL  
 30 GAEPLWAVGTDTPPSADWQPGRTTMGPGLWVVMGVLVGVAGGHDTYWTEQIDPWFLHGLG  
 LARTYWRDNTNGRLWLPNTPDASDPQGRGLAPPGELNLTASVPMLRWYAERFCFVLVTT  
 AEFPRDPGQLLYIPKTYLLGRPRNASLPPEAGPTSRPPAEVTQLKGLSHNPGASALLR  
 SRAWVTFAAAPDREGLTFPRGDDGATERHPDGRRNAPPPGPPAGTPRHPTTNLSIAHLHN  
 ASVTWLAARGLLRTPGR\*

35

Gene matched: gi|364588|prf||1508243Y  
 Gene name: UL22 gene [Human herpesvirus 1]

40

[SEQ ID NO:49]  
 ORF # = 12 from Contig 104  
 ORF start site = 21285  
 ORF end site = 20155  
 45 ORF sequence:

MASHAGQQHAPAFGQAARASGPTDGRAASRPSHRQGASEARGDPELPTLLRVYIDGPHGV  
 GKTTTSAQLMEALGPRDNIVYVPEPMTYWQVLGASETLTNIYNTQHRLDRGEISAGEAAV  
 VMTSAQITMSTPYAATDAVLAPHIGGEAVGPQAPPALTLVFDRHPIASLLCYPAARYLM  
 GSMTPOAVLAFVALMPPTAPGTNLVGLVPEAEHADRLARRQRPGERLDLAMLSAIRRVY  
 5 DLLANTVRYLQRGGRWREDWGRLTGVAATPRDPEDGAGSLPRIEDTLFALFRVPELLA  
 PNGDLYHIFAWVLDVLADRLLPMLHFLVDYDQSPVGCARDALLRLTAGMIPTRVTTAGSIA  
 EIRDLARTFAREVGGV\*

10 Gene matched: gi|59823  
 Gene name: (V00466) thymidine kinase [Herpes simplex virus ty

[SEQ ID NO:50]  
 15 ORF # = 13 from Contig 104  
 ORF start site = 20968  
 ORF end site = 22032  
 ORF sequence:  
 VLRVVDVRQGLGGPQHLPVSHRLGDVDDIVARPQGLHQLRGGGGLPHFVGSVYINPQQRG  
 20 QLRIPAGFGGFLAMARTGRRAAVGRPARTSSLTERRRVLLAGVRSHTRFYKAFAREVREF  
 NATRICGTLTLMMSGSLQGRSLFEATRVTLICEVDLGPRRPDCICVFEFANDKTLGGVCV  
 ILELKTCKSISSGDTASKREQRTTGKQLRHSCLKQLSLAPPGDKVVYLCPILVFVAQRT  
 LRVSRVTRLVPQKISGNITAAVRMLQSLSTYAVPPEPQTRRSRRRVAATARPQRPPSPTR  
 DPEGTAGHPAPPESDPPSPGVVGVAEEGGVQLKIAALFCVPVAAKSRPRTKTE\*

25  
 Gene matched: gi|136854|sp|P10208|UL24\_HSV11  
 Gene name: PROTEIN UL24. gi|74056|pir||

30  
 [SEQ ID NO:51]  
 ORF # = 14 from Contig 104  
 ORF start site = 22313  
 ORF end site = 23893  
 35 ORF sequence:  
 MDPYYPFDALDVWEHRRFIVADSRSFITPEFPRDFWMLPVFNI PRETAAERAAMQAQRT  
 AAAAALENAALQAAELPVDIERRIRPIEQQVHHIADALEALETAAAAEEADAARDAEAR  
 GEGAADGAAPSPTAGPAAAEVQIVRNDPPLRYDTNLVVDLLHVMYAGRGAAGSSGVVF  
 GTWYRTIQERTIADFPLTTRSADFRDGRMSKTFMTALVLSLQSCGRLYVGQRHYSAFECA  
 40 VLCLYLLYRTTHESSPDRDRAPVAFGDLLARLPYRLARLAAVIGDESGRPQYRYRDDKLP  
 KAQFAAAGGRYEHGALATHVVIATLVRHGVLPAPGDVPRDTSTRVNPDDVAHRDDVNRA  
 AAAFLARGHNLFLEWEDQTLRATANTITALAVLRLLANGNVYADRLDNRLQLGMLIPGA  
 VPAAEIARGASGLDSGAIKSGDNNLEALCVNYVLPYQADPTVELTQLFPGAGRVPVGRP  
 GGAATGVDEARGGYVVGRRPPGARAPHRAGAHQPHPHKHHPCGGDY\*

45



Gene matched: gi|136863|sp|P10209|UL25\_HSV11  
Gene name: VIRION PROTEIN UL25. gi|7406

5

[SEQ ID NO:52]

ORF # = 15 from Contig 104

ORF start site = 23784

ORF end site = 24071

10

ORF sequence:

VVDMLSGARQAALVRLTALELINRTRTNTTPVGEIINAHDALGIQYEQGLGLLAQQARIG  
LASNAKR FATFNVGSDYDLLYFLCLGFIPQYLSVA\*

15

Gene matched: gi|136863|sp|P10209|UL25\_HSV11

Gene name: VIRION PROTEIN UL25. gi|7406

[SEQ ID NO:53]

20

ORF # = 16 from Contig 104

ORF start site = 24292

ORF end site = 25638

ORF sequence:

25

VRVPMASAEMRERLEAPLPDRAVPIYVAGFLALYDSGDPGELALDPDTVRAALPPENPLP  
INVDHRARCEVGRVLAVVNDPRGPFVGLIACVQLERVLETAASAAIFERRGPALSREER  
LLYLITNYLPSVSLSTKRRGDEVPPDRTLFAHVALCAIGRRLGTIVTYDTSLDAAIAPFR  
HLDPATREGVRREAEEALALAGRTWAPGVEALTHLTLLSTAVNNMMLRDRWSLVAERRRQ  
AGIAGHTYLQASEKFKIWGAESAPAPERGYKTGAPGAMDTSPAASVPAPQVAVRARQVAS  
SSSSSSFPAPADMNPVSASGAPAPPPPGDGSYLWIPAFHYNQLVTGQSAPHHPPLTACG  
30 LPAAGTVAYGHPGAGPSPHYPPPPAHYPYGYAVRGPQSPGGPDRRAGGGHRRRPPGGWAS  
GRRRRPRDPGVGEPPPTRGGAAGVRLRP\*

35

Gene matched: gi|1224097

Gene name: (U49329) UL26 protease [Human herpesvirus 2]

[SEQ ID NO:54]

ORF # = 17 from Contig 104

40

ORF start site = 25463

ORF end site = 26221

ORF sequence:

45

MLFAGPSPLEAQIAALVGAIAADRQAGGLPAAAGDHGIRGSANRRRHEVEQPEYDCGRDE  
PDRDFPYYPGEARPEPRPVDSSRAARQASGPHETITALVGAVTSLQQELAHMRARTHAPY  
GPYPYPVGPYHHPHADTETPAQPFRYPAEAVYLPPPHIAPPGPPLSGAVPPPSYPPVAVTP

GPAPPLHQPSPAHAHPPPPPGPTPPPAASLPQPEAPGAEAGALVNASSAAHVNVDTARA  
ADLFVSQMMGSR\*

- 5 Gene matched: gi|1224097  
Gene name: (U49329) UL26 protease [Human herpesvirus 2]

- 10 [SEQ ID NO:55] = Contig ID 14

- [SEQ ID NO:56]  
ORF # = 1 from Contig 14  
15 ORF start site = 665  
ORF end site = 787  
ORF sequence:  
VKYQPRKLGKFKFNNLRDCGLYQRSPLQKFARLDIQPLLH\*

20

Gene matched: gi|76474|pir||JQ0950  
Gene name: ICP 18.5 protein - infectious laryngot

25

[SEQ ID NO:57] = Contig ID 38

- 30 [SEQ ID NO:58]  
ORF # = 1 from Contig 38  
ORF start site = 273  
ORF end site = 43  
ORF sequence:  
35 VELTAAQGVLPVSDSTSGDRAQLNNNNNNDDDDYNNKKQLKPLQTQTLSHSFEVSSGS  
PNTEVEIGERTDYLLK\*

- Gene matched: gi|1020200  
40 Gene name: (U31782) minor capsid protein L2 [Human papillom

- [SEQ ID NO:59] = Contig ID 50  
45

- [SEQ ID NO:60]  
ORF # = 1 from Contig 50  
ORF start site = 365  
ORF end site = 3  
5 ORF sequence:  
MSRRSPRRRGPRRRPRPGGPTVPRPGAFTADSQMVPAYDSGTAVESAPAASSLLRRWLL  
VPQADSDDDADYAGNDDAEWANSPPSEGGGKAPEAPHAAPASACPPPPPRKERGQQRPLP  
X
- 10 Gene matched: gi|132753|sp|P28283|RL1\_HSV2H  
Gene name: NEUROVIRULENCE FACTOR (ICP34).
- 15 [SEQ ID NO:61] = Contig ID 53
- [SEQ ID NO:62]  
ORF # = 1 from Contig 53  
20 ORF start site = 754  
ORF end site = 380  
ORF sequence:  
VETAHARMYPDAPPLRLCRGANVRYRVTRFGPDTLVPMSPREYRRAVLPAIDGRAASG  
AGDAMAPGAPDFCÉDEAHSHRACARWGLGAPLRPVYVALGRDTVRGGPADLLGPRREFCA  
25 RALL\*
- Gene matched: gi|124141|sp|P08392|ICP4\_HSV11  
Gene name: TRANS-ACTING TRANSCRIPTIONAL
- 30
- [SEQ ID NO:63] = Contig ID 67
- 35 [SEQ ID NO:64]  
ORF # = 1 from Contig 67  
ORF start site = 487  
ORF end site = 26  
ORF sequence:  
40 VSDGQHQAIVXXEVQASEPYIRVANGFGLVVPQGGQGTIDTXELHXDTNLDIRSGDEVHY  
HVTAGRRWGQLLWATQSVTAFSQEDLLDGAIFYRLNGSLRTRDTLIFSMEMGPVHTDATI  
QVTVALEGPLAPLKLVRHKKIYVFXGRGSWGIL\*
- 45 Gene matched: gi|560570|bbs|151525

Gene name: chondroitin sulfate proteoglycan NG2=t

[SEQ ID NO:65]

5 ORF # = 2 from Contig 67

ORF start site = 353

ORF end site = 511

ORF sequence:

VELXCVNGALTSLRDHKAKTVGHTDVGLRGLHLXXHSGLVLPIGHTCSWIQP\*

10

Gene matched: gi|1079684

Gene name: (U39205) Lpel4p [Saccharomyces cerevisiae]

15

[SEQ ID NO:66] = Contig ID 74

20 [SEQ ID NO:67]

ORF # = 1 from Contig 74

ORF start site = 224

ORF end site = 412

ORF sequence:

25 MITGLDNNVCYPITQFAIYNRLTCDKTYRIMPEYAHEAMNVFVNDQVYNWLCGSEIPFKY  
LK\*

Gene matched: gi|550075

30 Gene name: (D10935) cephalosporin-C deacetylase [Bacillus su

[SEQ ID NO:68] = Contig ID 76

35

[SEQ ID NO:69]

ORF # = 1 from Contig 76

ORF start site = 111

ORF end site = 1

40 ORF sequence:

MALTEDASSDSPTSAPKXTPLPVSATAMDQAYRYSXX

45 Gene matched: gi|138297|sp|P13290|VGLG\_HSV2

Gene name: GLYCOPROTEIN G. gi|419139|pir

[SEQ ID NO:70] = Contig ID 82

5

[SEQ ID NO:71]

ORF # = 1 from Contig 82

ORF start site = 767

10 ORF end site = 1156

ORF sequence:

VALAPYVNKTVTGDCLPVLDMETGHIGAYVVLVDQTNVADLLRAAAPAWSRRTLLPEHA  
RNCVRPPDYPTPPASEWNSLWMTFVGNMLFDQGTLVGALDFHGLRSRHPWSREQGAPAPA  
GDAPAGHGE\*

15

Gene matched: gi|124135|sp|P28284|ICP0\_HSV2H

Gene name: TRANS-ACTING TRANSCRIPTIONAL

20

[SEQ ID NO:72] = Contig ID 87

[SEQ ID NO:73]

25 ORF # = 1 from Contig 87

ORF start site = 519

ORF end site = 1475

ORF sequence:

30 MLNDMQWLASSDSEETEVEGISDDDLHRDSTSEAGSTDTEMFEAGLMDAATPPARPPAER  
QGSPTPADAQSGCGGPGVGEAAEAGGGDVCAVCTDEIAPPLRCQSFPCLHPFCIPCMK  
TWIPLRNTCPLCNTPVAYLIVGVTASGSFSTIPIVNDPRTRVEAAAVRSGTAVDFIWTG  
NPRTAPRSLSLGGHTVRALSPTPPWPGTDDEDDDPDGEGRGSGTGRGSGTGRGSGTGR  
GSGTGRGSGGGQALTGG SRLCLPLQPELISRPPNTSPPGA AVPGPPLVTPPPLLPNLRP  
PAPPGTTLTRGPPFLGRGF

35

Gene matched: gi|124135|sp|P28284|ICP0\_HSV2H

Gene name: TRANS-ACTING TRANSCRIPTIONAL

40

[SEQ ID NO:74] = Contig ID 89

45 [SEQ ID NO:75]

ORF # = 1 from Contig 89

ORF start site = 259

ORF end site = 615

ORF sequence:

5 MLADRWKHTDGNWYWFDSNGEMATGWKKIADKWYFNEEGAMKTGWVKYKDTWYYLNAK  
EGAMVSNAFIHSAGRNRLVLPQTRPNTGRQARIHSRPRWLDYVKIIMECLSNQNPAYY\*

Gene matched: gi|113676|sp|P06653|ALYS\_STRPN

10 Gene name: AUTOLYSIN (N-ACETYLMURAMOYL-

[SEQ ID NO:76] = Contig ID 90

15

[SEQ ID NO:77]

ORF # = 1 from Contig 90

ORF start site = 507

20 ORF end site = 2702

ORF sequence:

VKTIKSMDMPVATSFLAPDGTPLQYALCFPAVTDKLGALLMRPEAACVRPPLPTDVLESA  
PTVTAMYVLTVVNRLLQLALSDAQANFQLFGRFVRHRQATWGASMDAAAEYVALVATTL  
TREFGCRWAQLGWASGAAAPRPPPGPRGSQRHCVAFNENDVLVALVAGVPEHIYNFWRLD  
25 LVRQHEYMHLTLERAFEDAAESMLFVQRLTPHPDARIRVLPTFLDGGPPTRGLLFGTRLA  
DWRRGKLSETDPLAPWRSALGLGTQRRDAPALGKLSPAQALAAVSVLGRMCLPSAALAAL  
WTCMFPDDYTEYDSFDALLAARLESGQTLGPAGGREASLPEAPHALYRPTGQHVAVLAAA  
THRTPAARVTAMDVLAAVLLGAPVVVALRNTTAFSRESELELCLTLFDSRPGGPDAALR  
DVVSSDIETWAVGLLHTDLNPIENACLAQLPRLSALIAERPLADGPPCLVLVDISMTPV  
30 AVLWEAPEPPGPPDVRVFGSEATEELPFVATAGDVLAASAADADPFFARAILGRPFDASL  
LTGELFPGHPVYQRLADEAGPSAPTAARDPRDLAGDGGSGPEDPAAPPARQADPGVLA  
PTFLTATDTGEPVPPRMWAWIHGLEELASEDAGGPTPNPAPALLPPPATDQSVPTSQYAP  
RPIGPAXTARETRPSVPPQNTGRVPVAPRXDERPSPTPSPPADAAVPPPAFSGFAAAF  
SAAVPRVRRSRR

35

Gene matched: gi|135576|sp|P10220|TEGU\_HSV11

Gene name: LARGE TEGUMENT PROTEIN (VIRI

40

[SEQ ID NO:78] = Contig ID 91

[SEQ ID NO:79]

45 ORF # = 1 from Contig 91

ORF start site = 364

ORF end site = 2751

ORF sequence:

VCPPPTGATVVQFEQPRRCPTRPEGQNYTEGIAVVFKENIAPYKFKATMYKDVTVSQVW  
5 FGHYSQFMGIFEDRAPVPFEEVIDKINAKGVCSTAKYVRNNMETTAFHRDDHETDMEL  
KPAKVATRTSRGWHTTDLKYNPSRVEAFHRYGTTVNCIVEVDARSVYPYDEFVLATGDF  
VYMSPFYGYREGSHTETSIAADRFKQVDGFYARDLTTKARATSPTTRNLLTPKFTVAW  
DWVPKRPAVCTMTKWQEVDEMLRAEYGGSFSSDAISTTFTTNLTQYSLSRVDLGDCIG  
RDAREAIDRMFARKYNATHIKVGQPQYLLATGGFLIAYQPLLSNTLAELYVREYMRQDR  
10 KPRNATPAPLREAPSANASVERIKTTSSIEFARLQFTYNHIQRHVNDMLGRIAVAWCELO  
NHELTWNEARKLNPNAIASATVGRRVVSARMLGDVMAVSTCVVPAPDNVIVQNSMRVSSR  
PGTCYSRPLVSFRYEDQGPIEGQLGENNELRLTRDALEPCTVGHRRYFIFGGGYVYFEE  
YAYSHQLSRADVTTVSTFIDLNITMLEDFVPLEVYTRHEIKDSGLLDYTEVQRRNQLH  
DLRFADIDTVIRADANAAMFAGLCAFFEGMDLGRAVGKVMGVVGGVVSASVSGVSSFMS  
15 NPPGALAVGLLVLAGLVAAFFAFRYVLQLQRNPMKALYPLTTKELKTSDPGGVGGEEG  
AEGGGFDEAKLAEAREMIRYMXLVSAMERTEHKARKKGTSALLSSKVTNMVLRKRNKARY  
SPLHNEDEAGDEDEL\*

20 Gene matched: gi|138198|sp|P06763|VGLB\_HSV23

Gene name: GLYCOPROTEIN B PRECURSOR. gi

25

[SEQ ID NO:80] = Contig ID 93

[SEQ ID NO:81]

30 ORF # = 1 from Contig 93

ORF start site = 533

ORF end site = 1678

ORF sequence:

VALFVPLRLGWDPTGLVVRVERASWGPPAAPRAALLDVEAKVNFNPLALAAARVAEHPGA  
35 RLAWARLAAIRNSPQCASSASLAVTITTRTARFAREYTTLAFPPTSKEGAFADLVEVCEV  
CLRPRGHPHRVTARVLLPRGYNFYVSAGDGFSAPALVALFRQWHTTVHPAPGALAPVFAF  
LGPGEVRRGGPLQYFAVLGFGWPPFTVPAAAAAESVRDLLRGAACHTPLCPGGPGPRWA  
PRSSCPRGHGRPWPRRRPAASCPPFGKRWGGTTPRPPPSNYSTPRRPSGRSGRRGFVSPG  
SRPSSWPPSRASGRPGCRKPGGRAWKGWTRWWRPPRSPGPGPCWSAWCRTRATPAPRS  
40 GSCSAGSWPPSACRSSRRPAR\*

Gene matched: gi|136802|sp|P10192|UL08\_HSV11

Gene name: PROTEIN UL8. gi|73829|pir||w

45

[SEQ ID NO:82]  
ORF # = 2 from Contig 93  
ORF start site = 1288  
5 ORF end site = 2448  
ORF sequence:  
VASEAAGRLLPAFREAVARWHTATTIQLLDPPAAVGPVWTARFCFSGLQAQLLAALAGL  
GEAGLPEARGRAGLERLDALVAAAPSEPWARAVLERLVPDACDACPALRQLLGGVMAAVC  
LQIEQTASSVKFAVCGGTGAAFWGLFNVDPGDADAAHGAIHDARRALEASVRVLSANGI  
10 RPRLAPSLALEGVYTHVVTWSQTGAFFWNSRDDTDFLQGFPLRGPAYAAAAEVMRDALRR  
ILRRPAAGPPEEAACAARGIMEDACDRFVLDAFGRRLDAEYWSVLTTPPGEADDPLPQTAF  
RGGALLDAEQYWRVRVRCPPGGESVGPVVDLYPRPLVLPVDCAHHLREILREIQLVFT  
GVLEGVWGEGGSFVYPFEKMRFLFP\*

15  
Gene matched: gi|136802|sp|P10192|UL08\_HSV11  
Gene name: PROTEIN UL8. gi|73829|pir||w

20  
[SEQ ID NO:83]  
ORF # = 3 from Contig 93  
ORF start site = 3631  
ORF end site = 2705  
25 ORF sequence:  
VRRTRAGASNAGMADPTPADEGTAAAILKQAIAGDRSLVEVAEGISNQALLRMACEVRQV  
SDRQPRFTATSVLVDVTPRGRLRFVLDGSSDDAYVASEDYFKRCGDQPTYRGFAVVVLT  
ANEDHVHSLAVPPLVLLHRLSLFRPTDLRDFELVCLLMYLENCPRSHATPSLFVKVSAWL  
GVVARHASPFEVRCLLLRSCHWILNTLMCMAGVKPFDELVLPHWYMAHYLLANNPPPV  
30 LSALFCATPQSFAQLPGVPVRTDCVAYNPAGVMGSCWKSDDLRSALVYWWLSGSPKRRT  
SSLFYRFC\*

Gene matched: gi|136798|sp|P10191|UL07\_HSV11  
35 Gene name: PROTEIN UL7. gi|73828|pir||w

[SEQ ID NO:84]  
ORF # = 4 from Contig 93  
40 ORF start site = 4286  
ORF end site = 3570  
ORF sequence:  
MSPATQLQARDRELRRQAAGALEREHRAADRAAGGAGRPAEADLLRADYDIIIDVSKSMD  
DDTYVANSFQHQYIPAYGQDLERLSRLWEHELVRCKILRHRNNQCGQETSISYSSGAIAS  
45 FVAPYFEYVLRAPRAGALITGSDVILGEEELWEAVFKKTRLQTYLTDVAALFVADVQHAA



LPRPPSPTPADFRASASPRGGSRSRTRTRSRSRPGRTPRGAPDQGWGVQRRDGRPHARR\*

Gene matched: gi|136794|sp|P10190|UL06\_HSV11

5 Gene name: VIRION PROTEIN UL6. gi|73994

[SEQ ID NO:85] = Contig ID 94

10

[SEQ ID NO:86]

ORF # = 1 from Contig 94

ORF start site = 3669

ORF end site = 496

15 ORF sequence:

PRLSRAYLRHARGFEGSPGDTYPLRIGRRQSFPPGPAVSAPRRRARTPVAMSDSALQVPA  
PAGMTPPSAPPNGPLQVLLGSLTNLRRPPSPSSEEPAGSADEPAFLSAAKLRAATAAFL  
SGAAVGPAEARACWHPLLEQLCALHRAHGLPETALLAENLPGLLVHRMAVALPETPEAAF  
REMDVIKDTVLAITGSDTTHALEAAGLRRTTAALGPVVRVQCAVEWIDRWRTVTQSCLAMN  
20 PRTSLEALGEMSLKMSPVPLGQPGANLTPAYSLLFPSPIVQEGRLFLALVSNWVTLFSA  
HLQRIDDAALTPLTRALFTLALVDEYLTTPDRGAVVPPPLLAQFQHTVREIDPAIMIPPL  
EATKMVRSREEVRVSTALSRVSPRSACAPPGTLMARVRTDAAVFDPDPVFLSASALAIFR  
PAVTGLLQLGEPSSAGAQQRLALLQQTWALVQNSNSPSVVINTLTAGFTPAHCTQYIS  
ALEGFLVAGVPARTPPGHGLSEIQQLFGCIALAGANVFGLAREYGHYAGYVKTFRRIQGA  
25 SEHTHGRLC EAVGLSGGVLSQTLARIMGPAVPTEHLASLRRTL VGEFETAERRFSAGQPS  
LLRETALIWLVDYQGTHWDLTPPTPATPLSALLPVGPPSHAPSVHLAAATKIRFPALEGI  
HPNVLADPGFVPYVLALVVGDALRATCNAAYLPRPIEFALRVLAWARDFGLGYLPTVEGH  
RTKLGALITLLEPATRAGVGPTMQMADNIEQLRELYVIARGAVEQLRPAVQLPPPQPPE  
VGSSLLLISMYALAARGVMQEF AERADPLVRQLEDAIVLLRLHMRTLA AFFECRFESDGH  
30 RLYAVVADAHERLGPWRPEAMGDAVSQYCGMYHDAKRALVASLAGLSVVTETT AHLGVC  
DELAAQVSHEGNVLAVVRREIHGFLAIVSGIHARASKLMSGDQVPGFCYMSQFLARWRRL  
SAGYQAARAATGPERVAEFVQELHDTWKGLQTERALVVAPFASSGDQRTAAIQEVMAHAN  
EDAPPARPQTRRAHKRHDWGAGXTXXGAWVXDVVXDS\*

35

Gene matched: gi|221758

Gene name: (D10879) UL37 [Herpes simplex virus type 1]

40 [SEQ ID NO:87] = Contig ID 95

[SEQ ID NO:88]

ORF # = 1 from Contig 95

45 ORF start site = 371

ORF end site = 18

ORF sequence:

VLLDAPAPTASGRKTPAQGLAKEVQFSTAPPSPTAPWTPRVAGFNKRVFCAAVGRLAAT  
HARLAAVQLWDMSPHTDGLNELLDLTTRVTVCEGKNLLQRANELVNPDAAQGIL\*

5

Gene matched: gi|136927|sp|P10233|UL49\_HSV11

Gene name: TEGUMENT PROTEIN UL49. gi|73

10

[SEQ ID NO:89]

ORF # = 2 from Contig 95

ORF start site = 831

15 ORF end site = 436

ORF sequence:

MTSRRSVKSCPREAPRGTHEELYYGPVSPADPESPRDDFRRGAGPMRARPRGEVRFLHYD  
EAGYALYRDSSESSEDNDESRTARPRRSASVAGSHGPGPARAPPPGGPVGAGGRSHAPP  
ARTPKMTRGAP\*

20

Gene matched: gi|136927|sp|P10233|UL49\_HSV11

Gene name: TEGUMENT PROTEIN UL49. gi|73

25

[SEQ ID NO:90]

ORF # = 3 from Contig 95

ORF start site = 1441

ORF end site = 2550

ORF sequence:

30 MSQWGPRAILVQTDSTNRNADGDWQAQAAVAIRGGGVVQLNMVNKRAVDFTPAECGDSEWAV  
GRVSLGLRMAMPDFCAIIHAPAVSGPGPHVMLGLVHSGYRGTVLAVVVSPNGTRGFAPG  
ALRVDVTFDIRATPPTLTPEPSSLHRFPQLAPSPLAGLREDPWLDGALATAGGAVALPAR  
RRGGSILVYAGELTQVTTEHGDCVHEAPAFLEPKREEDAGFDILIHRAVTVPANGATVIQPS  
LRVLRAADGPEACYVLGRSSLNARGLLVMPTRWPSGHACAFVVCNLTGVPVTLQAGSKVA  
35 QLLVAGTHALPWIPPDNIHEDGAFRAYPRGVDPDATATPRDPPILVFTNEFDADAPPSKRG  
AGGFGSTGI\*

40

Gene matched: gi|118955|sp|P10234|DUT\_HSV11

Gene name: DEOXYURIDINE 5'-TRIPHOSPHATE

45

[SEQ ID NO:91]

ORF # = 4 from Contig 95

ORF start site = 3535

ORF end site = 2756

ORF sequence:

VGWKGAGAEPRACSGMASLLGVLCGWGTRPEEQYEMIRAADPPSEAEPRLQEALAVVNA  
5 LLPAPITLDDALESDDTRRLVKARALARTYHACMVNLERLARHHPGLEGSTIDGAVAAH  
RDKMRRLLADTCMATILQMYMSVGAADKSADVLVSQAIRSMAESDVVMEDVAIAERALGLS  
TSALAGGTRTAGLGATEAPPGPTRAQAPEVASVPVTHAGDRSPVRPGPVPPADPTPDPRH  
RTSAPKRQASSTEAPLLLA\*

10

Gene matched: gi|136933|sp|P10235|UL51\_HSV11

Gene name: PROTEIN UL51. gi|73813|pir||

15

[SEQ ID NO:92]

ORF # = 5 from Contig 95

ORF start site = 2889

ORF end site = 5042

20

ORF sequence:

VTGTDATSGACALVPGGASVAPSPAVRVPPARAEVERPRARSAIATSSMTTSLSAMLRM  
AWETSTSADLSAAPTDMYICRMVAMHVSARRRILSRCAATAPSMVEPSSPGWWRASLSRL  
TMQAWYVRARARAFTRRRVSSSDSRASSVMGAGKSALTARASC SRGSASEGGSAAARII  
SYCCSSGRVPQPHSTPSRDAIPEHARGSAPAFPHPTPSGFAGAMGTEDCDHEGRSVAAPV  
25 EVMALYATDGCVITSSLALLTNCLLGAEPYIFSYDAYRPDAPNGPTGAPTEQERFEGSR  
ALYRDAGGLNGDSFRVTFCLLGTEVGVTTHHPKGRTRPMFVCRFERADDDAVLQDALGRGT  
PLLPAHITATLDLEATFALHANIIMALTVAIVHNAPARIGSGSTAPLYEPGESMRVSVGR  
MSLGQRGLTTLFVHHKARVLAAYRRAYYGSAQSPFWFLSKFGPDKKSLVLAARYYLLQAP  
RLGGAGATYDLQAVKDICATYAI PHDPRPDTLSAASLTSAAITRFCCTSQYSRGAAAAG  
30 FPLYVERRIAADVRETGALEKFIHDRSCLRVSDREFITYIYLAHFECFSPPRLATHLRA  
VTTHDPSPAASTEQPSPLGREAVEQFFRHVRAQLNIREYVKQNVTPRETALAGDAAAAYL  
RARTYAPAALTAPAYCGVADSS TKMMGR LAEERLLVPHGWPAFAPTPGDDAGGGI

35

Gene matched: gi|136939|sp|P10236|UL52\_HSV11

Gene name: DNA REPLICATION PROTEIN UL52

40

[SEQ ID NO:93] = Contig ID 96

[SEQ ID NO:94]

ORF # = 1 from Contig 96

ORF start site = 2599

45

ORF end site = 1064

ORF sequence:

VGGCVDKLPLLKTPGPVARGARWLARATRRMACRKFCGVYRRPDKRQEASVPPETNTAPA  
FPASTFYTPAEDAYLAPGPPETIHPSRPPSPGEAARLCQLQEILAQMHSEDEYPIVDAAG  
AEEDEADDDAPDDVAYPEDYAEGRFLSMVSAAPLPGASGHPPVPGRAAPPDVRTCDSGK  
5 MGATGFTPEELDTMDREALRAISRCKPPSTLAKLVTGLGFAIHGTLIPGSEGCVFSSH  
PNYPHRVIVKAGWYASTNHEARLLRRLNHPAILPLDLHVVSQVTCVLVLPKYHCDLYTYL  
SKRPSPLGHLQITAVSRQLLSAIDYVHCEGIIHRDIKTENILINTPENICLGDFGAACFV  
RGRSSPFFHYGIAGTIDTNAPEVLGDPYTQVIDIWSAGLVIFETAVHTASLFSAPRDE  
RRPCDNQIARIIRQAQVHVDEFPTHAESRLTAHYRSRAAGNNRPAWTRPAWTRYKIHDT  
10 VEYLICKALTFDAALRPSAAELLRLPLFHPK\*

Gene matched: gi|125617|sp|P13287|KR1\_HSV2

Gene name: SERINE/THREONINE-PROTEIN KINAS

15

[SEQ ID NO:95]

ORF # = 2 from Contig 96

20 ORF start site = 2795

ORF end site = 3373

ORF sequence:

MGVVVSVVVTLLNQNALPRTSADASPALWSFLLRQCRILASEPLGTPVVVRPANLRLRLA  
EPLMDLPKFTRPIVTRSCRCPNNTTGLFAEDDPLESIEILDAPACFRLLHQERPGPHR  
25 LYHLWVVGAAADLCVPFLEYAQKTRLGFRFIAMKTNDAWGEPWPLPDRFLPERTVSWTPF  
PAAPNHPLGKSP\*

Gene matched: gi|137125|sp|P13292|US02\_HSV2

30 Gene name: PROTEIN US2. gi|419137|pir||A

[SEQ ID NO:96]

ORF # = 3 from Contig 96

35 ORF start site = 3534

ORF end site = 3671

ORF sequence:

MGRPEIPDEPSWQTGDDDPQNPGPPLAVGDEWPPSSHVCYPITNL\*

40

Gene matched: gi|137125|sp|P13292|US02\_HSV2

Gene name: PROTEIN US2. gi|419137|pir||A

45

[SEQ ID NO:97]

ORF # = 4 from Contig 96

ORF start site = 5400

ORF end site = 3853

5 ORF sequence:

VGRMRVGERERGGKKKKEGRRRRKREGGEGKGKEEGGEEGEVREKGERDRGGEGGGREK  
RGEKGDGGGPRSQHPRFIAGRAPPSTGHRCGNWRQGVATMADIPDPPIAVNTTPANHA  
PPSPPPGSRKRRRPVLPSSSESEGKPDTESESSSTESSEDEAGDLRGRRRSPRELGGRY  
FLDLAESTTGTESEGTGPSDDDDDDASDGLVDTPPRKSKRPRINLRLTSSPDRRAGVV  
10 FPEVWRNDRPIRAAQPAQAQSSGDRAAAPRRSARQAQMRSGAAWTLDLHYIRQCVNQLF  
RILRAAPNPPGSANRLRHLVRDCYLMGYCRTLGPRTWGRLLQISGGTWDVRLRNAREV  
EARFEPAAEPVCELPCLNARRYGPECVGNLETNGGSTSDDEISDATDSDDTLASHSDTE  
GGPSPAGRENPEASGGAIAARLECEFGTFDWTSEEGSQPWLSAVVADTSSAERSGLPAP  
GACRATEAPEREDGCRKMRFPACPYPCGHTFLRP\*

15

Gene matched: gi|124184|sp|P04485|IE68\_HSV11

Gene name: IMMEDIATE-EARLY PROTEIN IE68

20

[SEQ ID NO:98] = Contig ID 98

25 [SEQ ID NO:99]

ORF # = 1 from Contig 98

ORF start site = 612

ORF end site = 872

ORF sequence:

30 MVMAACPTEPPGSGVPADQPRVQSSRTWRPPLVNSRELYRAQRAARCASSSDTPQAPGW  
CGGTCRHAVFGVVAVVVVILAFLLWR\*

Gene matched: gi|136952|sp|P28282|UL56\_HSV2H

35 Gene name: PROTEIN UL56. gi|73833|pir||

[SEQ ID NO:100]

ORF # = 2 from Contig 98

40 ORF start site = 1689

ORF end site = 1045

ORF sequence:

MWGPGPARFIARPGTHGRRVFTDPPPRNMTTTPLSNLFRLAPDITHVAPPYCLNATWQAE  
NALHTTKTDPACLAARSYLVRASCSTSGPIHCFFFAVYKDSQHSPLVTELRFADLVNH  
45 PPVLRELEDKRGRLRCTGPFSCGTIKDVSGASPAGEYTINGIVYHCHCRYPFKTCWL

ASAALQHLLRFISSSGTAARAAEQRRHKIKIKIKV\*

Gene matched: gi|136947|sp|P28281|UL55\_HSV2H

5 Gene name: PROTEIN UL55. gi|73806|pir||

[SEQ ID NO:101]

ORF # = 3 from Contig 98

10 ORF start site = 2705

ORF end site = 1821

ORF sequence:

MALSLTPPHADGRAPVPERKAPSADTIDPAVRAVLRSISERAVERISESFGRSALVMQD

PFGGMPFPAANSPWAPVLATQAGGFDAETRRVSWETLVAHGPSLYRTFAANPRAASTAKA

15 MRDCVLRQENLIEALASADETLAWCKMCIHHNLPLRPQDPIIGTAAAVLENLATRLRPFL

QCYLKARGLCGLDDLCSSRRRLSDIKDIASFVLVILARLANRVERGVSEIDYTTVGVGAGE

TMHFYIPGACMAGLIEILDTHRQECSSRVCELTASHTIAPLYVHGKYFYCNSLF\*

20 Gene matched: gi|124181|sp|P28276|IE63\_HSV2H

Gene name: TRANSCRIPTIONAL REGULATOR IE

25 [SEQ ID NO:102]

ORF # = 4 from Contig 98

ORF start site = 4922

ORF end site = 3906

ORF sequence:

30 MLAVRSLQHLTTVIFITAYGLVLAWYIVFGASPLHRCIYAVRPAGAHNDTALVWMKINQT

LLFLGPPTAPPGGAWTPHAHVCIYANIEGRAVSLPAIPGAMSRRVMNVHEAVNCLEALWD

TQMRLVVVGWFLYLAFVALHQRRCMFGVVSPAHSMPAPATYLLNYAGRIVSSVFLQYPYT

KITRLLCELSVQRQTLVQLFEADPVTFLYHRPAVGVI VGCCELLLRFVALGLIVGTALISR

GACAITYPLFLTITTWCFVSIIALTELYFILRRDSAPKNAEPAAPRGRSKGWSGVCGRCC

35 SIILSGIAVRLCYIAVVAGVVLMAIRYEQEIQRRLFDL\*

Gene matched: gi|116105|sp|P22485|CELF\_HSV2H

Gene name: CELL FUSION PROTEIN PRECURSO

40

[SEQ ID NO:103]

ORF # = 5 from Contig 98

ORF start site = 6334

45 ORF end site = 4874

ORF sequence:

AAFDLEVPGHRPFAPGPALPPGGLAVGGHMYVNRNEIFNAALAVTNIILDLIALKEPVP  
 FPRLEALGHFMRGALAAVXLLFPAARVNPDAYPCYFFKSACRPRAPPVCAGDGPSAGGD  
 DGDGDWFPDAGGDDGDEEWEEDTDPMDTTHGPLPDDEAAYLDLLHEQIPAATPSEPSV  
 5 CSCADKIGLRVCLPVPAPYVHGS�TMRGVARVIQAVLLDRNFVEAVGSHVKNFLIDT  
 GUYAHGHSRLRPYFAKIGPDGSACGRLLPVFVIPPACEDVPAFVAHADPRRFHFHAPPM  
 FSAAPREIRVLHSLGGDYVSFFEKKASRNALEHFGRRETITEVLGRYDVRPDAGETVEGF  
 ASELLGRIVACIEAHFPEHAREYQAVSVRRRAVIKDDWLLQLIPGRGALNQSLSCLRFKH  
 GRASRATARTFLALSVGTNNRLCASLCQQCFATKCDNNRLHTLFTVDAGTPCSRSAPSST  
 10 SRPSSS\*

Gene matched: gi|136939|sp|P10236|UL52\_HSV11

Gene name: DNA REPLICATION PROTEIN UL52

15

[SEQ ID NO:104] = Contig ID 99

20 [SEQ ID NO:105]

ORF # = 1 from Contig 99

ORF start site = 213

ORF end site = 659

ORF sequence:

25 VGVGVRGWWGGXCGGLVVGWCWVXWGGWGVFFCFFLFCXXFXXXXXXXXFLAPDLTDPL  
 LFAYVGFQVVNHGLMFVVPDIAVYAMLGGAVWISLTQVLGLRRRLHKDPDAGPWAAATLR  
 GLFFSVYALGFAAGVLVRPRMAASRRSG\*

30 Gene matched: gi|807644

Gene name: (M10053) unknown protein [Herpes simplex virus ty

[SEQ ID NO:106]

ORF # = 2 from Contig 99

35 ORF start site = 757

ORF end site = 2403

ORF sequence:

MGAGVPWTGIKARGAGGPITVRVLGWEVAQKATHPCCSCPREAVVSGNPPRCAGRAHRSF  
 AGAGALLVMALGRVGLAVGLWGLLVGVVVVLANASPGRTITVGPRGESNAAPSASPRN  
 40 ASAPRTTPTPPQPRKATKSKASTAKPAPPPKTGPPKTSSEPVRCNRHNPLARYGLRVQIR  
 CRFPNSTRTESRLQIWRYATATDAEIGTAPSLSEVMVNVSAPPGGQLVYDSAPNRTDPHV  
 IWAEGAGPGASPRLYSVVGPLGRQRLIIIEELTLETQGMYYVWGRTRDRPSAYGTWVRVRV  
 FRPPSLTIHPHAVLEGQPFKATCTAATYYPGNRAEFVWFEDGRRVFDPAQIHTQTQENPD  
 GFSTVSTVTSAAVGGQGPRTFTCQLTWHRDSVSFSRRNASGTASVLPRPTITMEFTGDH  
 45 AVCTAGCVPEGVTFWFLGDDSSPAEKVAVASQTSCGRPGTATIRSTLPVSYEQTEYICR

LAGYPDGIPVLEHHGSHQPPPRDPTERQVIRAVEGAGIGVAVLVAVVLAGTAVVYLTHAS  
SVRYRRLR\*

- 5 Gene matched: gi|138220|sp|P06475|VGLC\_HSV23  
Gene name: GLYCOPROTEIN C PRECURSOR. gi

10

{SEQ ID NO:107}

ORF # = 3 from Contig 99

ORF start site = 2634

ORF end site = 3152

- 15 ORF sequence:

MAFRASGPAYQPLAPAASPARARVPAVAWIGVGAIVGAFALVAALVLVPPRSSWGLSPCD  
SGWQEFNAGCVAWDPTPVEHEQAVGGCSAPATLIPRAAAKHLAALTRVQAERSSGYWWVN  
GDGIRTCLRLVDSVSGIDEFCEELAIRICYYPSPGGFVRFVTSIRNALGLP\*

20

Gene matched: gi|136917|sp|P06483|UL45\_HSV23

Gene name: PROTEIN UL45 HOMOLOG (18 KD)

- 25 {SEQ ID NO:108}

ORF # = 4 from Contig 99

ORF start site = 4072

ORF end site = 3419

ORF sequence:

- 30 MAGAPPRLPPRNPAPPEQRPAAARPLAAHREAAGVYNAVRTWGPDAEAEPDQMENTYLL  
PEDDAAMPAGVGLGSTPAADTTAAAWPAESHAPRAPSEEADSIYESVSEDGGRVYEEIPW  
VRVYENICLRRQDAGGAAPPGDAPDSPYIEAENPLYDWGGSALFSPPGATRAPDPGLSLS  
PMPARPTNALANDGPTNVAALSALLTKLKRGRHQSH\*

35

Gene matched: gi|114350|sp|P10230|ATI2\_HSV11

Gene name: ALPHA TRANS-INDUCING FACTOR

- 40 {SEQ ID NO:109}

ORF # = 5 from Contig 99

ORF start site = 5584

ORF end site = 4391

ORF sequence:

- 45 MQRRARGASSLRLARCLTPANLIRGANAGVPERRIFAGCLLPTPEGLLSAAVGVLQRAD



DLQPAFLTGADRSVRLAARHHNTVPESLIVDGLASDPHYDYIRHYASAAKQALGEVELSG  
GQLSRAILAQYWKYLQTVVPSGLDIPDDPAGXCDPSLHVLMPRTLLPKLVVRAPFKSGAA  
AAKYAAAVAGLRDAAHRLQQYMFMRPADPSRPSTDTALRLSELLAYVSVLYHWASWMLW  
TADKYVCRRLLGPADRRFVALSGSLEAPAETFARHLDRGPSGTGSMQCMALRAAVSDVLG  
5 HLTRLAHLWETGKRSGGTYGIVDAIVSTVEVLSIVHHAQYIINATLTGYVWVASDSLNN  
EYLRAAVDSQERFCRTAAPLFPTMTAPSWARMELSIK\*

Gene matched: gi|114351|sp|P08314|ATI2\_HSV1F

10 Gene name: ALPHA TRANS-INDUCING FACTOR

[SEQ ID NO:110]

ORF # = 6 from Contig 99

15 ORF start site = 7758

ORF end site = 5668

ORF sequence:

MSVRGHAVRRRRRASTRSHAPSAHRAESPVEDEPEGGGVGLMGYLRAVFNVDDSEVEAAG  
EMASEEPPRRRRREARGHPGSRRASEARAAAPRRASFPRPRSVTARSQSVRGRDRAIT  
20 RAPRGGYLGPMDDPRDLGRVGGSRVVPSPFLDELSEEDDYPAVAHDDGAGARPPATV  
EIEGRVSGPELQAAFPLDRLTPRVAAWDESVRSAALGHPAGFYPCPDSAFGLSRVGVM  
HFASPADPKVFFRQTLQQGEALAWYVTGDAILDLTDRRAKTSPSRAMGFLVDAIVRVAIN  
GWVCGTRLHTEGRGSELDDRAAELRRQFASLTALRPVGAAAVPLLSAGGAAPPHPGPDAA  
VFRSSLGSLLYWPGVRALLGRDCRVAARYAGRMTYIATGALLARFNPGLVAVKCVLPREAAF  
25 AGRVLDVLAVLAEQTVQWLSVVVGARLPHSAHPAFADVEQEALFRALPLGSPGVVAAEH  
EALGDTAARRLLATSGLNAVLGAAVYALHTALATVTLKYALACGDARRRRDNAAAAARVL  
ATGLILQRLGLADTVVACVALAFAFDGGSTAPEVGTYTPLRYACVLRATQPLYARTTPAK  
FWADVRAAAEHVDLRPASSAPRAPVSGTADPAFLLEDLAAPPAPLNSESVLGPRVRVVD  
30 IMAQFRKLLMGDEETAALRAHVSGRRATGLGGPPRP\*

Gene matched: gi|136920|sp|P10231|UL47\_HSV11

Gene name: VIRION PROTEIN UL47 (82/81 K

35

[SEQ ID NO:111]

ORF # = 7 from Contig 99

40 ORF start site = 9949

ORF end site = 8279

ORF sequence:

VILKMRGGGREGMSVIGDARHPRQFPSQGPRPFSVAGPGSLPPSPPPGARALLIRLSKSL  
PDPTAPMDLLVNLFADADGVSPPPRPAGGPKNTPAAPPLYATGRLSQAQLMPSPMPV  
45 PPAALFNRLDDDLGFSAGPALCTMLDTWNEDLFSGFPTNADMYRECKFLSTLPSDVIDWG

DAHVPERSPIDIRAHGDVAFPTLPATRDELPSYIEAMAQFFRGELRAREESYRTVLANFC  
 SALYRYLRASVRQLHRQAHMRGRNRDLREMLRTTIADRYRETARLARVLFHLHLYLFLSR  
 EILWAAAYAEQMMRPDLFDGLCCDLESWRQLACLFQPLMFINGS�TVRGVPVEARRLRELN  
 HIREHLNLPLVRSAAAEPEGAPLTTPPVLQGNQARSSGYFMLLIRAKLDSYSSVATSEGE  
 5 SVMREHAYSRGRTRNNYGSTIEGLLDLPDDDDAPAEAGLVAPRMSFSLAGQRPRLSTTA  
 PITDVSIGDELRLDGEEVDMTPADALDDFDLEMLGDVESPSPGMTHDPVLYGALDVDDFE  
 FEQMFTDAMGIDDFGG\*

10 Gene matched: gi|1168549|sp|P29793|ATIN\_HSV23  
 Gene name: ALPHA TRANS-INDUCING PROTEI

15 TABLE 2

[SEQ ID NO:112] = Contig ID 1 .

20

[SEQ ID NO:113] >contigl (start 332 - stop 874)  
 MRTPADDVSWRYEAPSVIDYARIDGIFLRYHCPGLDTFLWDRHAQRAYLVNPFLFAAGFLEDLSHSVF  
 PA  
 DTQETTTRRALYKEIRDALGSRKQAVSHAPVRAGCVNFDYSRTTRCVGRRDLRPANTTSTWEPPVSSD  
 25 DE  
 ASSQSKPLATQPPVLALSNAAPRRVSPTRGRRRHTRLRRN\*

gi|136776|sp|P28278|VGLL\_HSV2H GLYCOPROTEIN L PRECURSOR

30

[SEQ ID NO:114] >contigl (start 747 - stop 1751)  
 MKRARSRSPPSRPSSPFRTPPHGGSPRREVGAGILASDATSHVCIASHPGSGAGYPTRLAAGSAVQ  
 RR  
 RPRGCPPGVMFSASTTPEQPLGLSGDATPPLPTSVPLDWAAFRRAFLIDDAWRPLLEPELANPLTARL  
 LA  
 35 EYDRRCQTEEVLPREDVFSWTRYCTPDDVRVVIIGQDPYHHPGQAHGLAFSVRADVPVPPSLRNVLA  
 AV  
 KNCYPDARMSGRGCLEKWARDGVLLLNTTLTVKRGAAASHSKLGWDRFVGGVVRRLAARRPGLVFMLW  
 GA  
 HAQNAIRPDPRQHYVLKFSHPSPLSKVFPFGTCQHFLAANRYLETRDIMPIDWSV\*

40

gi|137039|sp|P28275|UNG\_HSV2H URACIL-DNA GLYCOSYLASE

[SEQ ID NO:115] >contigl (start 1806 - stop 2507)  
 MVKSRVSYSRVMSGVGEERVPSAFTILASGWTFAPQNHDPGASPNTTPIESIAGTAPDAHVGPLDGE  
 45 PD

RDAISPLTSSVAGDPPGADGPYVTFDTLTMVSSIDELGRRQLTDTIRKDLRLSLAKFSIACHTKTSSFS  
 GT  
 AARQRKRGAPPQRTCVPRSNKSLQMFVLCRANAAQVREQLRAVIRSRKPRKYYTRSSDGRLCFAVPV  
 FV  
 5 HEFVSSEPMRLHRDNVMLSTEPD\*

gi|136782|sp|P28279|UL03\_HSV2H PROTEIN UL3

[SEQ ID NO:116] >contig1 (start 3312 - stop2707)  
 10 MGNPQTTIAYSLHHPRASLTSAIPDAAQVVHVFESEGTAVLTRGRARQDRLPRGGVVIQHTPIGLLVI  
 ID  
 CRAEFCAYRFIGRASTQRLERWWDAMHAYPPFDSWSSSHGESVRSATAGILTVVWTPDTIYITATIY  
 GT  
 APEAARGCDNAPLDVRPTTPPAPVSPTAGEFPANTDLLVEVLREIQISPTLDDADPTPGT\*

15 gi|136788|sp|P28280|UL04\_HSV2H PROTEIN UL4

[SEQ ID NO:117] >contig1 (contig1 start 6024 - stop 3379)  
 20 MAASGGEGSRDVRAPGPPQPGARPAVRFRDEAFNFTSMHGVQPIIARIRELSQQQLDVTQVPRLO  
 WF  
 RDVAALEVPTGLPLREFPFAAYLITGNAGSGKSTCVQTLNEVLDCVVTGATRIAAQNMVVKLSGAFLS  
 RP  
 INTIFHEFGFRGNHVQAQLGQHPYTLASSPASLEDLQRRDLTYWEVILDITKRALAAHGGEDARNEF  
 HA  
 25 LTALEQTLGLGQALTRLASVTHGALPAFTRSNIIIVIDEAGLLGRHLLTTVVYCWWMINALYHTPQYA  
 GR  
 LRPVLVCVGSPTQTASLESTFEHQKLRCVSRQSENVLTYLICNRTLREYTRLSHSWAIFINNKRCEVEH  
 EF  
 GNLMKVLEYGLPITEEHMQFVDRFVVPESYITNPANLPGWTRLFSSHKEVSAYMAKLHAYLKVTREGE  
 30 FV  
 VFTLPVLTFVSVKEFDEYRRLTQOPTLTMEKWITANASRITNYSQSQDQDAGHVRCEVHKKQLVVAR  
 ND  
 ITYVLNSQVAVTARLRKMVFGFDGTFRTFEAVLRDDSFVKTOGETSVEFAYRFLSRLMFGGLIHFYNF  
 LQ  
 35 RPGLDATQRTLAYGRLGELTAELLSLRDAAGASATRAADTSRSPGERAFNFKHLGPRDGGPDDFPD  
 DD  
 LDVIFAGLDEQQLDVIFYCHYALEEPETTAAVHAQFGLLKRAFLGRYLILRELFGFVFESAPFSTYVDN  
 VI  
 FRGCELLTGSPRGGLMSVALQTDNYTLMGYTYTRVFFAEELRRRHATAGVAEFLEESPLPYIVLRDQ  
 40 HG  
 FMSVVNTNISEFVESIDSTELAMAINADYGISSKLAMTITRSQGLSLDKVAICFTPGNLRNLNSAYVAM  
 SR  
 TTSSEFLHMNLNPLRERHERDDVISEHILSALRDPNVVIVY\*

45 gi|74000|pir||WMBEU5 gene UL5 protein - human herpesvirus 1

[SEQ ID NO:118] >contig1 (start 5594 - stop 7375) translated  
 MVLMSGRLRNAPESLTVMFCAAIRVAPVTTQSRSTSLRVCTHVLFDPDPALPVMRYAANGNSRSGRPVGT  
 KA  
 5 ATSRNHCRRGTCVTSSCCCESSRMRAMIGWTPCMDVKFKNASSLNRTAGLAPGCCGGPGARTSREPS  
 PP  
 DAAMAAQRARAPAMRTRGGDAALCAPEDGWVKVHPTPGTMLFREILLGQMGYTEGQGVYNVVRSEAA  
 TR  
 QLQAAIFHALLNATTYRDLEEDWRRHVVARGLQPQRLVRRYRNAREGDIAGVAERVFDTWRCTLR TTL  
 10 LD  
 FAHGVVDCFAPGGPSGPTSF PKYIDWLTCLGLVPILRKTREREATQRLGAF LRQHTLPRQLATVAGAA  
 ER  
 AGPGLLELAVAFDSTRMAEYDRVHIYYNHRRGEWLVRDPVSGQRGECLVLCPPPLWTGDRLVFDSPVQR  
 LC  
 15 PEIVACHALREHAHICRLNTASVKVLLGRKSDSERGVAGAARVVNKALGEDDETKAGSAASRLVRLI  
 IN  
 MKGMRHVGDINDTVRAYLDEAGGHLIDTPAVDHTLPGFGKGGTGRGSAAQDPGAR PQQLRQAFQTAVV  
 NN  
 INGMLEGYINNLFGTIERLRETNAGLATQLQARV  
 20  
 gi|136794|sp|P10190|UL06\_HSV11 VIRION PROTEIN UL6

25 [SEQ ID NO:119] = Contig ID 10 Length: 21036 Type: N Check:  
 7835 ..

[SEQ ID NO:120] >contig10 (start 5688 - stop 1) translated  
 VAGAAHMI PAALPHPTMKRQGD RDIVVTGVRNQFATDLEPGGSVSCMRSSLSFLSLLFDVGP RDVLSA  
 30 EA  
 IEGCLVEGGEWTRAAAGSGPPRMCSIIELPNFLEYPAARGGLRCVFSRVYGEVGFGEPTAGLLETQC  
 PA  
 HTFFAGPWAMRPLSYTLTIGPLGMGLYRDGTAYLFDPHGLPAGTPAFIAKVRAGDVYPYLTYAHD  
 RP  
 35 KVRWAGAMVFFVPSGPGAVAPADLTAAALHLYGASETYLQDEPFVERRVAITHPLRGEIGGLGALFVG  
 VV  
 PRGDGEGSGPVVPALPAPTHVQT PRADRPPEAPRGASGPPNTPQAGHPNRPPDDVWAAALEGTPPAKP  
 SA  
 PDAAASGPPHAAPPPQTPAGDAAEEAEDLRVLEVGAVPVGRHRARYSTGLPKRRRP TWTTPSSVEDLT  
 40 SG  
 ERPAPKAPPKAKKKSAPKKKAPVAAEVPASSPTPIAATVPPAPDT PPQSGQGGDDGPASPSSPSVL  
 ET  
 LGARRPPEPPGADLAQLFEVHPNVAATAVRLAARDAALAREVAACSQLTINALRSPYPAHPGLELCV  
 IF

FFERVLAFLENGARTHTQAGVAGPAAALLDFTLRMLPRKTAVGDFLASTRMSLADVAHRPLIQHVL  
 DE  
 NSQIGRLALAKLVLVARDVIRETDAFYGDLADLDLQLRAAPPANLYARLGEWLLERSRAHPNTLFAPA  
 TP  
 5 THPEPLHRIQALAQFARGEEMRVEAEAREMREALDALARGVDSVSQRAGPLTVMPVPAAPGAGGRAP  
 CP  
 PALGPEAIQARLEDVRIQARRAIESAIKEYFHRGAVYSAKALQASDSHDCRFHVASAAVPMVQLLES  
 LP  
 AFDQHTRDVAQRAALPPPPPLATSPQAILLRDLLQRCQTLDAPEDLAAWLSVLTDAATQGLIERKPLE  
 10 EL  
 ARSIHGINDQQARRSSGLAELQRFDALDAALAQQLDSDAAFVPATGPAPYVDGGGLSPEATRMAEDAL  
 RQ  
 ARAMEAAKMTAELAPEARSRLRERAHAEAMLNDARERAKVAHDAREKFLHKLQGVLRPLPDFVGLKA  
 CP  
 15 AVLATLRASLPAGWTDLADAVRGPPPEVTAALRADLWLLGQYREALEHPTPDATALAGLHPAFVVV  
 LK  
 TLFADAPETPVLVQFFSDHAPTIKAVSNAINAGSAAVATASPAATVDAAVRAHGALADAVSALGAAA  
 RD  
 PASPLSFLAALADSAAGYVKATRLALEARGAIDELTTLGSAADLVVQARRACAQPEGDHAALIDAAA  
 20 RA  
 TTAARESLAGHEAGFGGLLHAEGTAGDHSPSGRALQELGKVIGATRRRADELEAAVADLTAKMAAQRA  
 RG  
 SSERWAAGVEAALDRVENRAEFDVVELRRLQALAGTHGYNPRDFRKRAEQALAANA EAVTLALDTAFA  
 FN  
 25 PYTPENQRHPMLPPLAAIHLGWSAAFHAAAETYADMFRVDAEPLARLLRIAEGLLEMAQAGDGFI  
 DY  
 AVGRLADDMTSVPGLRRYVPPFQHGADYVELRDLDAIRADVHRALGGVPLDLAAAAEQISAARN  
 DP  
 EA  
 TAEIVRTGVTLPSPEDALVACAAALERVQSPVKNTAYAEYVAFVTRQDTAETKDAVVRAKQQR  
 AEA  
 30 TE  
 RVMAGLREALAARERRAQIEAEGLANLKTMLKVAVPATVAKTLDQARSVAEIAEQVEVLLDQTEK  
 TR  
 EL  
 DVPAVIWLEHAQRTFETHPLSAARGDGPGLARHAGRLGALFDTRRRVDALRRSLEEA  
 EAEWDEVWGR  
 FG  
 35 RVRGGAWKSPEGFRAMHEQLRALQDTTNTVSGLRAQPAYERLSARYQGVLGAKGAERAE  
 AVEEIGARV  
 TK  
 HTALCARLRDEVVRRVPWEMNFDALGRLLAEFDAAAADLAPWAVEEFRGARELIQYRMGLY  
 SAYARAG  
 GQ  
 TXXXXX  
 40  
 gi|135576|sp|P10220|TEGU\_HSV11 LARGE TEGUMENT PROTEIN  
  
 [SEQ ID NO:121] >contig10 (start 9322 - stop 5978) translated  
 MSDSALQVPAPAGMTPPSAPPPNGPLQVLLGSLTNLRRPPSPSSEPA  
 GSADEPAFLSAAKLAATAAF  
 45 LL

SGAAVGPAEARACWHPLLEQLCALHRAHGLPETALLAENLPGLLVHRMAVALPETPEAAFREMDVIKD  
 TV  
 LAITGSDTTHALEAAGLRRTAALGPVRVRQCAVEWIDRWRTVTQSCLAMNPRTSLEALGEMSLKMSPV  
 PL  
 5 GQPGANLTPAYSLLFPSPIVQEGRLFLALVSNWVTLFSAHLQRIDDAALPLTRALFTLALVDEYLT  
 TP  
 DRGAVVPPPLLAQFQHTVREIDPAIMIPPLEATKMVRSREEVRVSTALSRVSPRSACAPPGTLMARVR  
 TD  
 AAVFDPDPVFLSASALAIFRPAVTGLLQLGEPSSAGAQQRLLALLQQTWALVQNSNPSVINTLTDA  
 10 GF  
 TPAHCTQYISALEGFLVAGVPARTPPGHGLSEIQQLFGCIALAGANVFGLAREYGHYAGYVKTFRRIQ  
 GA  
 SEHTHGRLEAVGLSGGVLSQTLARIMGPAVPTEHLASLRRTLVEGEFETAERFSAGQPSLLRETALI  
 WL  
 15 DVYQGTWDLTPTTPATPLSALLPVGPPSHAPSVHLAAATKIRFPALEGIHPNVLADPGFVFPYLALV  
 VG  
 DALRATCNAAYLPRPIEFALRVLAWARDFGLGYLPTVEGHRTKLGALITLLEPATRAGVGPTMQMADN  
 IE  
 QLLRELYVIARGAVEQLRPAVQLPPPQPPEVGSSLLISMYALAARGVLQELAERADPLVRQLEDAIV  
 20 LL  
 RLHMRTLAAFFECRFESDGHRLYAVVADAHERLGPWRPEAMGDAVSQYCGMYHDAKRALVASLAGLRS  
 VV  
 TETTAHLGVCDELAQAQVSHEGNVLAVVRREIHGFLAIVSGIHARASKLMSGDQVPGFCYMSQFLARWR  
 RL  
 25 SAGYQAARAATGPERVAEFVQELHDTWKGLQTERALVVAPFASSADQRTAAIQEVMAHATEDAPPSPA  
 AD  
 LVVLTNRHDLGAWGDYSLGPLGQPTVVPDSVDLSPQGLAATLSMDWLLINELLQVTDGVFRASAFRPS  
 AG  
 PEAPGDLEAQDAGGSTPEPTTPGPQDTQARAPSTRPAGRETVPWPNTPVEDDEMTPOETPPVHP\*  
 30 gi|221758 UL37 [Herpes simplex virus type 1]  
  
 [SEQ ID NO:122] >contig10 (start 9262 - stop 11211) translated  
 VERTGGSCRRAPGPGARCPTWRPACALGDAARRPRAQTGMTAAALYGGAKYRPGTLRNPGRVASTPRR  
 35 RG  
 VLYGALCPGIPFVSGSGPAGVWECVCVGGGRRDGGPDQVYRGRSVGRPNRPFKHLRMHRPSQSDTGTH  
 QR  
 RKPPSPVRVRVFGGVFFLSALLPPLHHPPTTRPLAIGGKTMKTKPLPTAPMAWAESAVETTTSPR  
 EL  
 40 AGHAPLRRVLRPPIARRDGPVLLGDRAPRRTASTMWLLGIDPAESSPGTRATRDDTEQAVDKILRGAR  
 RA  
 GGLTVPGAPRYHLTRQVTLTDLCPNAERAGALLLALRHPTDLPHLARHRAPPGRQTERLAEAWGQLL  
 EA  
 SALGSGRAESGCARAGLVSFNFLVAACAAAYDARDAEAIVRAHITTNYGGTRAGARLDRFSECLRAMV  
 45 HT

HVFPHEVMRFFGGLVSWVTQDELASVTAVCSGPQEATHTGHPGRPCSAVTIPACAFVDLDAELCLGGP  
GA  
AFLYLVTYRQCRDQELCCVYVVKSQLPPRGLEAALERLFGRLRITNTIHGAEDMTPPPPNRNVDFPL  
AV  
5 LAASSQSPRCSASQVTNPQFVDRLYRWQPDLRGRPTARTCTYAAFAELGVMPDNSPRCLHRTERFGAV  
GV  
PVVILEGVVWRPGGWRACA\*

gi|139176|sp|P22486|VP19\_HSV2G CAPSID ASSEMBLY AND DNA MATU  
10 [SEQ ID NO:123] >contig10 (start 11673 - stop 15215) translated  
VIRRPVRPFGRTAHPASHGPAAVSVHRVRATVTLVPMANRPAASALAGARSPSERQEPREPEVAPPGG  
DH  
VFCRKVSGVMVLSSDPPGPAAYRISDSSFVQCGSNCSMIIDGDVARGHLRDLEGATSTGAFVAISNVA  
15 AG  
GDGRTAVVALGGTSGPSATTSVGTQTSGEFLHGNPRTPEPQGPQAVPPPPPPPPFWGHECCARRDARG  
GA  
EKDVGA AESWSDGPSSDSETEDSDSSEDTGSGSETLSRSSSIWAAGATDDDDSDSDSRSDSVQPDV  
VV  
20 RRRWSDGPAPVAFPKPRRPGDSPGNPGLGAGTGPGSATDPRASADSDSAHAAAPQADVAPVLD SQPT  
VG  
TDPGYFPVPLELTPENAEAVARFLGDAVDREPALMLEYFCRCAREESKRVP PRTFGSAPRLTEDDFGLL  
NY  
ALAEMRRLCLDLPVPPNAYTPYHLREYATRLVNGFKPLVRRSARLYRILGILVHLRIRTREASFEEW  
25 MR  
SKEVDLDFGLTERLREHEAQLMILAQALNPYDCLIHSTPNTLVERGLQSALKYEEFYLRFGGHYMES  
VF  
QMYTRIAGFLACRATRGMRHIALGRQGSWWEMFKFFHRLYDHQIVPSTPAMNLGTRNYTSSCYLV  
NP  
30 QATTNQATLRAITGNVSAILARNGGIGLCMQAFNDASPGTASIMPALKVLDLVAAHNKQSTRPTGAC  
VY  
LEPWHS D VRAVLRMKGVLAGEEAQRCDNIFSALWMPDLFFKRLIRHLDGEKNVTWSLFD RDTSM SLAD  
FH  
GEEFEKLYEHLEAMGFGETIPIQDLAYAIVRSAATTGSPFIMFKDAVNRHYIYDTQGA AIAGSNLCTE  
35 IV  
HPSSKRSSGVCNLGSVNLARCVRRTFDGMLRDAVQACVLMVNIMIDSTLQPTPQCARGHDNLRSMG  
IG  
MQGLHTACLKMGDLDES AEFRLNTHIAEVMLLAAMKTSNALCVRGARPF SHFKRSMYRAGR FHWERF  
SN  
40 ASPRYEGEWEMLRQSMMKHGLRNSQFIALMPTAASAQISDVSEGFAPLFTNLFSKVTRDGETLRPNTL  
LL  
KELERTFGGKRLLDAMDGLEAKQWSVAQALPCLDPAHPLRRFKTAFDYDQELLIDL CADRAPYVDHSQ  
SM  
TLYVTEKADGTL PASTLVRLLVHAYKRGLKTGMYCKVRKATNSGVFAGDDNIVCTSCAL\*  
45

gi|1710385|sp|P09853|RIR1\_HSV23 RIBONUCLEOSIDE-DIPHOSPHATE

[SEQ ID NO:124] >contig10 (start 15268 - stop 16281) translated  
MDPAVSPASTDPLDTHASGAGAAPIVPCPTPERYFYTSQCPDINHLRSL SILNRWLET ELVFVGDEED  
5 VS  
KLSEGELGFYRFLFAFLSAADDLV TENLGGLSGLFEQKDILHYVVEQECIEVVHSRVYNI IQLVLFHN  
ND  
QARRAYVARTINHPAIRVKVDWLEARVRECDSIPEK FILMILIEGVFFAASFAAIAYLR TNNLLRVTC  
QS  
10 ND LISRDEAVHTTASCIYNNYLGGHAKPEAA RVYRLFREAVDIEIGFIRSQA PTDS SILSPGALAAI  
EN  
YVRF SADRLGLIHMQLYSAPAPDASFPLSLMSTDKHTNFFECRSTSYAGAVVNDL\*

gi|132624|sp|P03174|RIR2\_HSV23 RIBONUCLEOSIDE-DIPHOSPHATE R

15 [SEQ ID NO:125] >contig10 (start 17637 - stop 16564) translated  
MRRRGHAFAPGDRGTRAAGPGPAAPWGAPSKPALRLAHLFCIRVLRALGYAYINSGQLEADDACANLY  
HT  
NTVAYVHTTDTDLLMGCDIVLDISTGYIPTIHCRDLLQYFKMSYPQFLALFVRCHTDLHPNNTYASV  
20 ED  
VLRECHWTAPSR SQARRAARRERANSRSLESMP TLTAAPVGLETRISWTEILAQQIAGEDDYEDPPL  
QP  
PDVAGGPRD GARSSSEILTPELVQVPNAQRVAEHRGYVAGRRRHVIHDAPEALDWLPDPMTIAELV  
EH  
25 RYVKYVISLISPKERGPWTLKRLPIYQDLRDEDLARSIVTRHITAPDIADRFLAQLWAHAPPPAFYK  
DV  
LAKFWDE\*

gi|549322|sp|P36699|VHS\_HSV2G VIRION HOST SHUTOFF PROTEIN

30 [SEQ ID NO:126] >contig10 (start 18537 - stop 19949) translated  
MAHLPGGAAAAPLSEDAIPSPRERTEDWPPCQIVLQGAELNGILQAFAPLRTSLLD SLLVVGDRGILV  
HN  
AIFGEQVFLPLDHSQFSRYRWGGPTAAFLSLVDQKRSLLSVFRANQYPDLRRVELTVTGQAPFRTL VQ  
35 RI  
WTTASDGEAVELASE TLMKRELTSFAVLLPQGD PDVQLRLTKPQLTKVVNAVGD ETAKPTTFELGPNG  
KF  
SVFNARTCVTFAAREEGASSSTSAQVQILTSALKKAGQAAANAKTVYGENTHRTFSVVVD DCSMRAVL  
RR  
40 LQVGGGTLKFFLTADVPSVCVTATGPN AVSAVFLKQPQRVCLNWLGRTPGSSTGSLASQDSRAGPTDS  
QD  
FSSEPDAGDRGAPEEEGLEQARVPPAFPEPPG TKRRHAGAEVVPAD DATKRPKTGVPAAPTRAESP  
LS  
ARYGPEAAEGGGDGGRYACYFRDLQTGDASPSPLSAFRGPQRPPYGFGLP\*

45



gi|136905|sp|P10226|VPAP\_HSV11 DNA POLYMERASE PROCESSIVITY

[SEQ ID NO:127] >contig10 (start 20031 - stop 21053) translated  
VCPPPTNMAVVGSGRLRLRPFHPPSPSFFVLRLALIRAGPGPFAASPRAPSGPGCGMCRGDSFPGVAGG  
5 SG  
EHCLGGDDGDDGRPRLACVGAIARGFAHLWLQATTLGFGVGSVLSRGPYADAMSGAFVIGSTGLGFLR  
AP  
PAFARPPTRVCAWLRLVGGGAVALWSLGEAGAPPVPGPATQCLALGAAYAALLVLADDVHPLFLLA  
PR  
10 PLFVGTGLGVVVGGLTIGGSARYWWIDPRAAAALTAAVVAGLGTAAAGDSFSKACPRHRRFCVVS  
PP  
PRYAPEDAERPTDHGPLLSTHHQSRPRVCGDGAARPENIWVPVVTFAGALALAACAARG

15 gi|136909|sp|P10227|UL43\_HSV11 MEMBRANE PROTEIN UL43

[SEQ ID NO:128] = Contig 11 Length: 2343 Type: N Check: 6656 ..  
[SEQ ID NO:129] >contig11 (start 2357 - stop 3) translated  
APLLVDLRALDARARASSPEGHEVDPQLRRRGEAYLRAGGDPGPLVLREAVSALDLPFATSFLAPD  
20 GT  
PLQYALCFPAVTDKLGALLMRPEAACVRPPLPTDVLESAPTVMYVLTVVNRLQLALSDAQANFQL  
FG  
RFVRHRQATWGASMDAAAEYVALVATTLTREFGCRWAQLGWASGAAPRPPGPRGSQRHCVAFNEN  
DV  
25 LVALVAGVPEHIYNFWRLDLVRQHEYMHLTLERAFEDAAESMLFVQRLTPHPDARIRVLPFLDGGPP  
TR  
GLLFGTRLADWRRGKLSETDPLAPWRSALGLTQRRDAPALGKLSPAQALAAVSVLGRMCLPSAALAA  
LW  
TCMFDDYTEYDSFDALLAARLESGQTLGPAGGREASLPEAPHALYRPTGQHVAVLAAATHRTPAARV  
30 TA  
MDLVLAAVLLGAPVVVALRNTTAFSRESELELCLTLFDSRPGGPDAALRDVVSSDIETWAVGLLHTDL  
NP  
IENACLAQLPRLSALIAERPLADGPPCLVLVDISMTPVAVLWEAPEPPGPPDVRVFGSEATEELPFV  
AT  
35 AGDVLAASAADADPFFARAILGRPFDSALLTGELFPGHPVYQRPLADEAGPSAPTAARDPRDLAGGDG  
GS  
GPEDPAAPPARQADPGVLAPTLTDTATGEPVPPRMWAWIHGLEELASDDAGGPTPNPAPALLPPPAT  
DQ  
SVPTSQYAPRPIGPAATARETRPSVPPQNTGRVPVAPRDDPRSPPTPSPPADAALPPPAFSGSAAA  
40 FS  
AAVPRVRRSRXXXXX

45 gi|135576|sp|P10220|TEGU\_HSV11 LARGE TEGUMENT PROTEIN

[SEQ ID NO:130] = Contig 12 Length: 14928 Type: N Check: 1371

..

[SEQ ID NO:131] >contig12 (start 1505 - stop 3) translated

5 MAAAPPAAVSEPTAARQKLLALLGQVQTYVFQLELLRRCDPQIGLGKLAQLKLNALQVRVLRRLRPG  
LE  
AQAAAFLTPLSVTLELLLEYAWREGERLLGHLETTFATTGDSAFFTETMGLARPCPYHQIRLETYGG  
DV  
RMELCFLHDVENFLKQLNYCHLITPPSGATAALERVREFMVAAVGSGLIVPPELSDPSHPCAVCFEEL  
10 CV  
TANQGATIARRLADRICNHVTQQAQVRLDANELRRYLPHAAGLSDAARARALCVLDQALARTAAGGGA  
RA  
GPPPADSSSVREEADALLEAHDFVQATTPGLYAISELRFWLASGDRARHSTMDAFADNLNALAQRELQ  
QE  
15 TAAVAVELALFGRRAEHFDRAFGGHLAALDMVDALIIGGQATSPDDQIEALIRACYDHHLTTPLLRRL  
VS  
PEQCDEEALRRVLARLGAGGATGGAEEEEPRAAAEEGRRRGAGTPASEDGERGPEPGAQGPESWGDI  
AT  
RAAADVXXXXX

20

gi|124088|sp|P10212|P RTP\_HSV111 PROCESSING AND TRANSPORT PRO

[SEQ ID NO:132] >contig12 (start 5468 - stop 1878) translated

25 MDTKPKTTTTVKVPPGPMGYVYGRACPAEGLELLSLLSARSGDADVAVAPLIVGLTVESGFANVA  
VG  
SRTTGLGGTAVSLKLMPSHYSPSVYVFHGGRHLPSTQAPNLTRLCERARRHFGFSDYAPRPCDLKHE  
TT  
GDALCERLGLDPDRALLYLVITEGFREAVCISNTFLHLGMDKVTIGDAEVHRI PVYPLQMFMPDFSR  
VI  
30 ADPFNCNHRSIGENFNYPFPFNRLARLLFEAVVGPAVALRARNVDAVARAAHLAFDENHEGAAL  
PA  
DITTFEASQGKPGRGARDAGNKGPAGGFEQRLASVMAGDAALALESI VSMVDFEPPPDITTWPLL  
EG  
QETPAARAGAVGAYLARAAGLVGAMVFTNSALHLTEVDDAGPADPKDHSKPSFYRFFLVPGTHVAAN  
35 PQ  
LDREGHVVPGYEGRPTAPLVGGTQEFAGEHLAMLCGFSPALLAKMLFYLERCDGGVIVGRQEMDVFRY  
VA  
DSGQTDVPCNLCTFETRHACAHTTLMRLRARHPKFASAARGAIGVFGTMNSAYSDCDVLGNAAFSAL  
KR  
40 ADGSENTRTIMQETYRAATERVMAELEALQYVDQAVPTALGRLETIIGTREALHTVVNNIKQLVDREV  
EQ  
LMRNLI EGRNFKFRDGLAEANHAMSLSLDPYTCGPCPLLQLLARRSNLAVYQDLALSQCHGVFAGQSV  
EG  
RNFRNQFQPVLRRLRRVMDLFNNGFLSAKTLTVALSEGA AICAPSLTAGQTAPAESSFEGDVARTLGFP  
45 KE

LRVKSRLVFAGASANASEAAKARVASLQSA YQKPKRVDILLGPLGFLKQFHAVIFPNGKPPGSNQF  
NP  
QWFWTALQRNQLPARLLSREDIETIAFIKRFSLDYGAINFINLAPNNVSELAMYYMANQILRYCDHST  
YF  
5 INTLTAVIAGSRRPPGVQAAAAWAPQGGAGLEAGARALMDSLDAHPGAWTSMFASCNLLRPVMAARPM  
VV  
LGLSISKYYGMAGNDRVFQAGNWASLLGGKNACPLLI FDRTRK FVLACPRAGFVCAASSLGGGAHEHS  
LC  
EQLRGIIAEGGA AVASSVFVATVKSLGPRTQQIQIEDWLALLEDEYLSEEMMEFTTRALERGHGEWST  
10 DA  
ALEVAHEAEALVSQ LGAAGEVFNFGDFGDEDDHAASFGGLAAAAGAAGVARKRAFHGDDPFGE GPPEK  
KD  
LTL DML\*

15 gi|544182|sp|P36384|DNBI\_HSV2 MAJOR DNA-BINDING PROTEIN

[SEQ ID NO:133] >contig12 (start 6286 - stop 10008) translated  
MFCAAGGPTSPGGKSAARAASGFFAPHNPRGATQTAPPPCRRQNFYNPHLAQTGTQPKAPGPAQRHTY  
YS  
20 ECDEFRIAPRSLDEDAPAEQRTGVHDGRLRRAPKVYCGGDERDVLRVGPEGFWPRRLRLWG GADHAP  
EG  
FDPTVTVFHVYDILEHVEHAYSMRAAQLHERFMDAITPAGTVITLLGLTPEGHRVAVHVYGRQYFYM  
NK  
AEVDRHLQCRAPRDL CERLAAALRESPGASFRGISADHFEAEVVERADVYYETRPTLYRVFVRSGR  
25 AL  
AYLCDNFCPAIRKYEGGV DATTRFILDNPGFVTFGWYRLKPGRGNAPAPRPPTAFGTSSDVEFNCTA  
DN  
LAVEGAMCDLPAYKLMCFDIECKAGGEDELAFVPAERPEDLVIQISCLLYDLSTTALEHILLFSLGSC  
DL  
30 PESHLSDLASRGLPAPVVLEFDSEFEMLLAFMTFVKQYGPEFVTGYNIINFDWPFVLTKLTEIYKVPL  
DG  
YGRMNGRGVFRVWDIGQSHFQKR SKIKVNGMVNIDMYGIIITDKVKLSSYKLNAVAEAVLKDKKKDLSY  
RD  
IPAYYASGPAQRGVIGEYCVQDSLLVGQLFFKFLPHLELSAVARLAGINITRTIYDQQIRVFTCLLR  
35 LA  
GQKGFILPDTQGRFRGLDKEAPKRPAVPRGEGERP GDGNGDEDKDDDEDGDEDEREEVARETGGRH  
VG  
YQGARVLDPTSGFHVDPVVVDFASLYPSIIQAHNLCFSTLSLRPEAVAHLEADRDYLEIEVGGRRLF  
FV  
40 KAHVRESLLSILLRDWLAMRKQIRSRIPQSTPEEAVLLDKQQA AIKVV CNSVYGFTGVQHGLLPCLHV  
AA  
TVTTIGREMLLATRAYVHARWAEFDQLLADFPEAAGMRAPGPYSMRIIYGDTDSIFVLCRGLTAAGLV  
AM  
GDKMASHISRALFLPPIKLECEKTF TKLLLI AKKKYIGVICGGKMLIKGVDLVRKNNCAFINRTSRAL  
45 VD

LLFYDDTVSGAAAALAERPAAEWLARPLPEGLQAFGAVLVDAHRRITDPERDIQDFVLTAELSRHPRA  
YT  
NKRLAHLTVYYKLMARRAQVPSIKDRIPYVIVAQTREVEETVARLAALRELDAAAPGDEPAPPAALPS  
PA  
5 KRPRETPSHADPPGGASKPRKLLVSELAEDPGYAIARGVPLNTDYYFSHLLGAACVTFKALFGNNAKI  
TE  
SLLKRFI PETWHPDDVAARLRAAGFGPAGAGATAEETRMLHRAFDTLA\*

gi|118882|sp|P07918|DPOL\_HSV21 DNA POLYMERASE

10 [SEQ ID NO:134] >contig12 (start 10870 - stop 9953) translated  
MYDIAPRRSGSRPGGRDKTRRRSRFSAAGNPGVERRASRKSLPSHARRLELCLHERRRYRGFFAALA  
QT  
PSEEIAIVRSLSVPLVKTTTPVSLPFSLDQTVADNCLTSLSGMGYYLGIGGCCPACSAGDGRLATVSREA  
15 LI  
LAFVQQINTIFEHRTFLASLVVLADRHSTPLQDLLADTLGQPELFFVHTILRGGGACDPRFLFYPDPT  
YG  
GHMLYVIFPGTSAHLHYRLIDRMLTACPGYRFAAHVWQSTFVLVVRRNAEKPADAEIPTVSAADIYCK  
MR  
20 DISFDGGLMLEYQRLYATFDEFPPP\*

gi|136875|sp|P10215|UL31\_HSV11 PROTEIN UL31

[SEQ ID NO:135] >contig12 (start 12674 - stop 10863) translated

25 VRPARPAMATSAPGVPSAAVREESPGSSWKEGAFERPYVAFDPDLLALNEALCAELLAACHVVGVP  
AS  
ALDEDVESDVAPAPPRPGAAREASGGRGPGSARGPPADPTAEGLLDTGPFAAASVDTFALDRPCLVC  
RT  
IELYKQAYRLSPQWADYAFCLCAKCLGAPHCAASIFVAAFEFVYVMDHHFLRTKKATLVGSFARFALT  
30 IN  
DIHRHFFLHCCFRTDGGVPGRHAQKQPRPTSPGAQKVQYSNYSFLAQSATRALIGTLASGGDDGAGA  
GG  
GSGTQPSLTALMNWKDCARLLDCTEGKRGGDSCCTRAAARNGEFEAAAGALAQQGEPETWAYADLI  
LL  
35 LLAGTPAVWESGPRLRAAADARRAAVSESWEAHRGARMRDAAPRFAQFAEPKAQPDLDLGPLMATVLK  
HG  
RGRGRTGGECLLCNLLLVRAYWLMRRLRASVVRYSENNTSLFDCIVPVVDQLEADPEAQPGDGGRFV  
SL  
LRAAGPEAIFKHMFCDFMCAITEMEVDPPWVLFHGPRADHRDELQLHKAKLACGNEFEGRVCIALRALI  
40 YT  
FKTYQVFVPKPTALATFVREAGALLRRHSISLLSLEHTLCTYV\*

gi|136879|sp|P10216|UL32\_HSV11 PROBABLE MAJOR ENVELOPE GLYC

45 [SEQ ID NO:136] >contig12 (start 12652 - stop 13044) translated

MAGRAGRTRPRTLRLDAIPDCALRSQTLES LDARYVSRDGAGDAAVWFEDMTPAELEVIFPTTDAKLN  
LS  
RTQRLASLLTYAGPIKAPDGPAPHTQDTACVHGELDATERERFAAVINRFLDLHQILRG\*

5 gi|136883|sp|P10217|UL33\_HSV11 PROTEIN UL33

[SEQ ID NO:137] >contig12 (start 13134 - stop 13964) translated  
MAGMGKPYGGRPGDAFEGLVQIRILIVPTTLRGGGGESGPYSPSNPPSRCAFQFHGQDGSDEAFPIEY  
VL  
10 RLMNDWADVPCNPYLRVQNTGVS VLFQGGFFNRPHGAPGGAITAEQTNVILHSTETTGLSLGDLDDVKG  
RL  
GLDARPMMAWMWISCFVRMPRVQLAFRFMGPEDAVRTRRILCRAAEQALARRRRSRSSQDDYGAVAVA  
AA  
HHSSGAPGPGVAASGPPAPPGRGPARPWHQAVQLFRAPRPGPPALLLLVAGLFLGAIIWWAVGARL\*

15 gi|136888|sp|P10218|UL34\_HSV11 VIRION PROTEIN UL34

[SEQ ID NO:138] >contig12 (start 14076 - stop 14414) translated  
MAAPQFHRPSTITADNVRALGMRGLVLATNNAQFIMDNSYPHPHGTQGAVREFLRGQAAALTDLGVTH  
20 AN  
NTFAPQPMFAGDAAAEWLRPSFGLKRTYSPFVVRDPKTPSTP\*

gi|139196|sp|P10219|VP26\_HSV11 CAPSID PROTEIN VP26

25 [SEQ ID NO:139] = Contig ID 13 Length: 838 Type: N Check: 7960  
..

[SEQ ID NO:140] >contig13 (start stop 852 - 1) translated  
30 RRLYADRLTKRSLASLGRCVREQRGELEKMLRVSVHGEVLPATFAAVANGFAARARFCALTAGAGTVI  
DN  
RAAPGVFDAHRFMRASLLRHQVDPALLPSITHRFFELVNGPLFDHSTHSFAQPPNTALYYSVENVGLL  
PH  
LKEELARFIMGAGGSGADWAVSEFQKFYCFDGVSGITPTQRAAWRYIRELIIATTLFASVYRCGELEL  
35 RR  
PDCSRPTSEGLYRYPPGVYLTYNSDCPLVAIVESGPDGCIGPRSVVVYDRDVFSLYSVLQHLAPRLA  
GXXXXX

40 gi|124089|sp|P12835|PRTP\_HSV1A PROCESSING AND TRANSPORT PRO

[SEQ ID NO:141] = Contig ID 14 Length: 2647 Type: N Check:  
2951 ..

45

[SEQ ID NO:142] >contig14 (start 2661 - stop 97) translated  
PPVSPATTKARKRKTKKPKRPEATPPPDANATVAAGHATLRAHLREIKVENADAQFYVCPPTGAT  
VV  
QFEQPRRCPTRPEGQNYTEGIAVVFKENIAPYKFKATMYKDVTVSQVWFGHRYSQFMGIFEDRAPVP  
5 FE  
EVIDKINAKGVCRSTAKYVRNNMETTAFHRDDHETMELKPAKVATRTSRGWHTTDLKYNPSRVEAFH  
RY  
GTTVNCIVEVDARSVYPYDEFVLATGDFVYMSPFYGYREGSHEHTSYAADRFKQVDGFYARDLTTK  
AR  
10 ATSPTRNLLTTPKFTVAWDWVPKRPAVCTMTKWQEVDMLRAEYGGSRFSSDAISTTFTNLQYS  
LS  
RVDLGDCIGRDAREAIDRMFARKYNATHIKVGQPQYYLATGGFLIAYQPLLSNTLAELYVREYMRQD  
RK  
PRNATPAPLREAPSANASVERIKTTSSIEFARLQFTYNHIQRHVNDMLGRIAVAWCELQNHETLWNE  
15 AR  
KLNPNAIASATVGRRV SARMLGDVMAVSTCVVPAPDNVIVQNSMRVSSRPGTCYSRPLVSFRYEDQGP  
LI  
EGQLGENNELRLTRDALEPCTVGHRRYFIFGGGYVYFEEYAYSHQLSRADVTTVSTFIDLNITMLEDH  
EF  
20 VPLEVYTRHEIKDSGLLDYTEVQRRNQLHDLRFADIDTVIRADANAAMFAGLCAFFEGMGDLGRAVGK  
VV  
MGVVGGVVS AVSGVSSFMSNPF GALAVGLLVLAGLVAAFFAFRYVLQLQRNPMKALYPLTTKELKTS  
PG  
GVGGE GEEGAEGGGFDEAKLAEAREMIRYMALVSAMERTEHKARKKGTSALLSSKVTNMVLRKRNKAR  
25 YS  
PLHNEDEAGDEDEL\*

gi|138198|sp|P06763|VGLB\_HSV23 GLYCOPROTEIN B PRECURSOR

30 [SEQ ID NO:143] = Contig ID 15 Length: 20389 Type: N Check: 2794  
..

[SEQ ID NO:144] >contig15 (start 788 - stop 3) translated  
35 MNAHFANEVQYDLTRDPSSPASLIHVIISSSECLAAAGVPLSALVRGRPDGGAAANFRVETQTRAHATG  
DC  
TPWRSFAAAYVPADAVGAILAPVIPAHPDLLPRVPSAGGLFVSLPVACDAQGVYDPYTVAALRLAWGP  
WA  
TCARVLLFSYDELVPPNTRYAADGARLMRLCRHFCRYVARLGAAAPAAATEAAHLSLGMGESGTPTP  
40 QA  
SSVSGGAGPAVVGTPDPPI SPEEQLTAPGGDTATAEDVSITQENEEIXXXXX

gi|136835|sp|P10201|UL17\_HSV11 PROTEIN UL17

45 [SEQ ID NO:145] >contig15 (start 818 - stop 2089) translated

VPEGAWVGGACARPRGPRAHVRLYAVCFVCPQGIRGQDFNLLFVDEANFIRPDAVQTIMGFLNQANCK  
 II  
 FVSSNTNGKASTSFLYNLRGADELNNVVTYICDDHMPRVVTHTNATACSCYILNKPVFITMDGAVRR  
 TA  
 5 DLFLPDSFMQEIIIGGQARETGDDRPVLTKSAGERFLLYRPSTTTNSGLMAPELYVYVDPAPTANTRAS  
 GT  
 GIAVVGRIYRDDFIIFALEHFFLRALTGSAPADIARCVVHSLAQVLALHPGAFRSVRVAVEGNSSQDSA  
 VA  
 IATHVHTEMHRILASAGANGPGPELLFYHCEPPGAVLYPFFLLNKQKTPAFEYFIKKFNSSGGVMASQ  
 10 EL  
 VSVTVRLQTDPVEYLSEQLNNLIETVSPNTDVRMYSGKRNGAADDLMVAVIMAIYLAAPTGIPTAFFP  
 IT  
 RTS\*

15 gi|139646|sp|P04295|VTER\_HSV11 PROBABLE DNA PACKAGING PROTE

[SEQ ID NO:146] >contig15 (start 3520 - stop 2429) translated  
 VLLSPAPPPLPHGRCPPSLFHHRPGCVALS GPPAPPRSGVSRPGAMITDCFEADIAIPSGISRPDAAA  
 LQ  
 20 RCEGRVFLPTIRRLALADVAHESFVSGGVSPDTLGLLLAYRRRFPVITRVLPTTRIVACPVDLGLT  
 HA  
 GTVNLNRTSPVDLCNGDPVSLVPPVFEGQATDVRLES LDTLRFPVPLPTPLAREIVARLVARGIRDL  
 NP  
 DPRTPGELPDNLVLYNGARLSLVADVQQLASVNTSLRSLVLMVYSITEGTTLILTLPRLALSAQ  
 25 DG  
 YVNALLQMOSVTREAAQLIHPEAPMLMQDGERRLPYEA LVAWLAHAGQLGDILALAPAVRVCTFDGA  
 AV  
 VQSGDMPVIRYP\*

30 gi|139191|sp|P10202|VP23\_HSV11 CAPSID PROTEIN VP23

[SEQ ID NO:147] >contig15 (start 7954 - stop 3764) translated  
 VWEGLGLPELGLMEPANPPRNPMAPARDPPGYRYAAAMVPTGSILSTIEVASHRRLFDF FARVRSDE  
 NS  
 35 LYDVEFDALLGSYCNTLSLVRFLGLSVACVCTKFPELAYMNEGRVQFEVHQPLIARDGPHPVQPV  
 HN  
 YMTKVIDRRALNAAFSLATEAIALLTGEALDGTGISLHRQLRAIQQLARNVQAVLGA FERGTADQMLH  
 VL  
 LEKAPPLALLLPMQRYLDNGRLATRVARATLVAELKRSFCDTSF FLGKAGHRREAIEAWLVDLTTATQ  
 40 PS  
 VAVPRLTHADTRGRPVDGVLVTAAIKQRLQLQSFLKVEDTEADVPTYGEMVLNGANLVTALVMGKAV  
 RS  
 LDDVGRHLLMQEEQLEANRETLELESAPQTTRVRADLVAIGDRLVFLEALEKRIYAATNPYPPLVG  
 AM

DLTFVLPLGLFNPAMERFAAHAGDLVPAPGHPEPRAFPQRQLFFWGKDHQVLRRLSMENAVGTVCHPSL  
 MN  
 IDAAVGGVNHDPVEAANFYGAYVAAPAGPGADMQQRFNLAWRQRLAHGRVRWVAECQMTAEQFMQPDN  
 AN  
 5 LALELHPAFDFFAGVADVELPGGEVPPAGPGAIQATWRVVNGNLPLALCPVAFRDARGLELGVGRHAM  
 AP  
 ATIAAVRGAFEDRSYPVFFYLLQAAIHGSEHVFCALARLVTQCITSYWNNTCAAFVNDYSLVSYIVT  
 YL  
 GGDLPPEECMAVYRDLVAHVEALAQLVDDFTLPGPELGGQAQAEINHLMRDPALLPPLVWDCDGLMRHA  
 10 AL  
 DRHRDCRIDAGGHEPVYAAACNVATADFNRNDGRLLHNTQARAADAADDRPHRPADWTVHHKIYYVYL  
 VP  
 AFSRGRCTAGVRFDRVYATLQNMVPEIAPGEECPSPDPTDPAHPLHPANLVANTVNAMEFHNGRVVV  
 DG  
 15 PAMLTQLVLAHNMAERTTALLCSAAPDAGANTASTANMRIFDGAHAGVLLMAPQHLDHTIQNGEYFY  
 VL  
 PVHALFAGADHVANAPNFPALRDLARHVPLVPPALGANYFSSIRQPVVQHARESAAGENALTYALMA  
 GY  
 FKMSPVALYHQLKTGLHPGFQFTVVRQDRFVTENVLFSEASEAYFLGQLQVARHETGGGVSFLLTQTP  
 20 RG  
 NVDLGVGYTAVAAATATVRNPVTDMGNLQNFYLGARGAPLLDNAAAVYLRNAVAVAGNRLGPAQLPVF  
 GC  
 AQVPRRAGMDHGQDAVCEFIATPVATDINYFRRPCNPRGRAAGGVYAGDKEGDVIALMYDHGQSDPAR  
 PF  
 25 AATANPWASQRFSGDLLYNGAYHLNGASPVLSPCFKFFTAADITAKHRCLERLIVETGSAVSTATAA  
 SD  
 VQFKRPPGCRELVEDPCGLFQEAYPITCASDPALLRSARDGEAHARETHFTQYLIYDASPLKGLSL\*

gi|137571|sp|P06491|VCAP\_HSV11 MAJOR CAPSID PROTEIN (MCP)

30 [SEQ ID NO:148] >contig15 (start 8869 - stop 8201) translated  
 MTMRDDVPLLDRELVEAACGGEDGELPLDEQFSLSSYGTSDFFVSSAYSRLPPHTQPVFVSKRVVMFA  
 WS  
 FLVLKPLELVAAGMYGWTGRAVAPACIIAAVLAYYVTWTLARALLLVNKRDLPLSPPVFWGLCVI  
 35 MG  
 GAALCALVAAAHETFSPDGLFHWITASQLLPRTDPLRARS LGIACAAGAAMWVAADCFAAFTNFFLA  
 RF  
 WTRAILKAPVAF\*

40 gi|136841|sp|P10204|UL20\_HSV11 MEMBRANE PROTEIN UL20

[SEQ ID NO:149] >contig15 (start 9205 - stop 11118) translated  
 VGRQGERWVGGGNEENTQRATSGMRPELSLKGRPCVTEAVVCPSTDAAIHSGGSSSVRPQPYARAARA  
 RA



THGSRSRHRQPLPPSSHHPTIPPPSPPRGSPAMELSYATTLHHRDVVFYVTADNRNAYFVCGGSV  
 YS  
 VGRPRDSQPGEIAKFLVVRGTGPKDRMVANYVRSELRQGLRDVRPVGEDEVFLDSVCLLNPNVSSE  
 RD  
 5 VINTNDVEVLDECLAEYCTSLRTSPGVLVTGVRVRARDRVIELFEHPAIVNISSRFAYTPSPYVFALA  
 QA  
 HLPRLPSSLEPLVSGFLDGIAPAPRQPLDARDRTDVITGTRAPRPMAGTGAGGAGAKRATVSEFVQV  
 KH  
 IDRVVSPSVSSAPPPSAPDASLPPPGQLQEAAPPGPPLRELWWVFYAGDRALEEPHAESGLTREEVRAV  
 10 HG  
 FREQAWKLFSGVAPRAFLGAALALSPTQKLAVYYYLIHRERRMSFPFALVRLVGRYIQRHGLYVPAP  
 DE  
 PTLADAMNGLFRDALAAGTVAEQLLMFDLLEPKDVPVGSADARADSAALLRFVDSQRLTPGGSVSPEHV  
 MY  
 15 LGAFGLVLYAGHGRLAAATHARTLTGVTSLVLTVGDVDRMSAFDRGPAGAAGRTRTAGYLDALLTVCL  
 AR  
 AQHGQSV\*

gi|136845|sp|P10205|UL21\_HSV11 PROTEIN UL21

20 [SEQ ID NO:150] >contig 15 (start 14107 - stop 11339) translated  
 VSISAGVRGQGWHRISTPPKNGAGRSVLVFGVLVPLCFYPHPTPSFGPRLRQQRASDSLGAEPLWAV  
 GT  
 DTPPSADWQPGRTTMGPGLWVMGVLVGVAGGHDTYWTEQIDPWFLHGLGLARTYWRDNTNTRLWLPN  
 25 TP  
 DASDPQRGRLAPPGELNLTASVPMRWYAERFCFVLVTTAEFPRDPGQLLYIPKTYLLGRPRNASLP  
 EL  
 PEAGPTSRPPAEVTQLKGLSHNPGASALLRSRAWVTFAAAPDREGLTFPRGDDGATERHPDGRNAPP  
 PG  
 30 PPAGTPRHPTTNLSIAHLHNASVTWLAARGLLRTPGRYVYLSPSASTWPVGWTTGGLAFGCDAALVR  
 AR  
 YGKGFMGLVISMRSPPAEIIVVPADKTLARVGNPTDENAPAVLPGPPAGPRYRVFVLGAPTPADNGS  
 AL  
 DALRRVAGYPEESTNYAQYMSRAYAEFLGEDPGSGTDARPSLFWRLAGLLASSGFAFVNAHAHAIR  
 35 LS  
 DLLGFLAHSRVLAGLAARGAAGCAADSVFLNVSVDPAARLRLEARLGHVAAILEREQSLAAHALGY  
 QL  
 AFVLDSPAAYGAVAPSAARLIDALYAEFLGGRALTAPMVRRALFYATAVLRAPFLAGAPSAEQRERAR  
 RG  
 40 LLITTALCTSDVAAATHADLRAALARTDHQKNLFWLPDHFSPCAASLRFDLAEGGFILDALAMATRS  
 IP  
 ADVMAQQTRGVASVLRWAHYNALIRAFVPEATHQCSGSPSHNAEPRILVPITHNASYVVTHTPLRGI  
 GY  
 KLTGVDVRRPLFITYLATCEGHAREIEPKRLVRTENRRDLGLVGAVFLRYTPAGEVMSVLLVDTDAT  
 45 QQ

QLAQGPVAGTPNVFSSDVPSVALLLFPNGTVIHLLAFDTLPATIAAGFLAASALGVVMITAALAGIL  
RV  
VRTCVPFWRRE\*

- 5 gi|138315|sp|P06477|VGLH\_HSV11 GLYCOPROTEIN H PRECURSOR
- [SEQ ID NO:151] >contig15 (start 15322 - stop 14192) translated  
MASHAGQQHAPAFGQAARASGPTDGRAASRPSHRQGASEARGDPELPTLLRVYIDGPHGVGKTTTSAQ  
LM  
10 EALGPRDNIVYVPEPMTYWQLGASETLTNIYNTQHRLDRGEISAGEAAVMTSAQITMSTPYAATDA  
VL  
APHIGGEAVGPQAPPALTLVFDRHPIASLLCYPAARYLMGSMTPOAVLAFVALMPPTAPGTNLVLGV  
LP  
EAEHADRLARRQRPGERLDLAMLSAIRRVYDLLANTVRYLQRGGRWREDWGRLTGVAATPRPDPEDG  
15 AG  
SLPRIEDTLFALFRVPELLAPNGDLYHIFAWVLDVLADRLLPMHLFVLDYDQSPVGCRDALLRLTAGM  
IP  
TRVTTAGSIAEIRDLARTFAREVGGV\*
- 20 gi|125438|sp|P04407|KITH\_HSV23 THYMIDINE KINASE
- [SEQ ID NO:152] >contig15 (start 15005 - stop 16069) translated  
VLRVVDVRQGLGGPQHLPVSHRLGDVDDIVARPQGLHQLRGGGGLPHVPGSVYINPQQRGQLRIPAGF  
GG  
25 PLAMARTGRRAAVGRPARTSSLTERRRVLLAGVRSHTRFYKAFAREVREFNATRICGTLTLMMSGSLQ  
GR  
SLFEATRVTLICEVDLGPFRPDCICVFEFANDKTLGGVCVILELKTCKSISSGDTASKREQRTTGMKQ  
LR  
HSLKLLQSLAPPGDKVVYLCPILVFVAQRTLRSRVTRLVPQKISGNITAAVRMLQSLSTYAVPPEPQ  
30 TR  
RSRRRVAATARPQRPPSPTRDPEGTAGHPAPPESDPPSPGVGVAAEGGGVLQKIAALFCVPVAAKSR  
PR  
TKTE\*
- 35 gi|136854|sp|P10208|UL24\_HSV11 PROTEIN UL24
- [SEQ ID NO:153] >contig15 (start 16350 - stop 18107) translated  
MDPYYPFDALDVWEHRRFIVADSRSFITPEFPRDFWMLPVFNIPRETA AERA AVLQAQRTAAAAALEN  
AA  
40 LQAAELPVDIERRIRPIEQVHHIADALEALETAAAAEEADAARDAEARGE GAADGAAPSPTAGPAA  
AE  
MEVQIVRNDPPLRYDTNLPVDLLHVMYAGRGAAGSSGVVFGTWYRTIQERTIADFPLTTRSADFRDGR  
MS  
KTFMTALVLSLQSCGRLYVGQRHYSAFECAVLCLYLLYRTTHESSPDRDRAPVAFGDLLARLPYRLAR  
45 LA

AVIGDESGRPQYRYRDDKLPKAQFAAAGGRYEHGALATHVVIATLVRHGVLPAA PGDVPRDTSTRVNP  
 DD  
 VAHRDDVNRAAAFLARGHNFLWEDQTL LRATANTITALAVLRLLANGNVYADRLDNRLQLGMLIP  
 GA  
 5 VPAAEIARGASGLDSGAIKSGDNNLEALCVNYVLPYQADPTVELTQLFPGLAALCLDAQAGRPLAST  
 RR  
 VVDMSSGARQAALVRLTALELINRTRTNTTPVGEI INAHDALGIQYEQGLGLLAQQARIGLASNAKRF  
 AT  
 FNVGSDYDLLYFLCLGFIPQYLSVA\*  
 10 gi|136863|sp|P10209|UL25\_HSV11 VIRION PROTEIN UL25  
  
 {SEQ ID NO:154} >contig15 (start 18328 - stop 20256) translated  
 VRVPMASAE MRERLEAPLPDRAVPIYVAGFLALYDSGDPGELALDPDTVRAALPPENPLPINVDHRAR  
 15 CE  
 VGRVLAVVNDPRGPFFVGLIACVQLERVLETAASAAIFERRGPALSREERLLYLITNYLPSVSLSTKR  
 RG  
 DEVPPDRTLFAHVALCAIGRRLGTIVTYDTSLDAAIAPFRHLDPATREGVRREAAEALAGRTWAP  
 GV  
 20 EALTHTLLSTAVNNMLRDRWSLVAERRRQAGIAGHTYLQASEKFKIWGAESAPAPERGYKTGAPGAM  
 DT  
 SPAASVPAPQVAVRARQVASSSSSSSFAPADMNPVSASGAPAPPPPGDGSYLWIPAFHYNQLVTGQ  
 SA  
 PHHPPLTACGLPAAGTVAYGHPGAGPSPHYPPPPAHYPGMLFAGPSPLEAQIAALVGAIADRQAGG  
 25 LP  
 AAAGDHGIRGSAKRRRHEVEQPEYDCGRDEPDRDFPYYPGEARPEPRPVDSRRAARQASGPHETITAL  
 VG  
 AVTSLQQLAHMRARTHAPYGPYPVGPYHHPHADTETPAQPPRYPAEAVYLPPPHIAPPGPPLSGAV  
 PP  
 30 PSYPPVAVTPGPAPPLHQPSPAHAHPPPPPGPTPPPAASLPQPEAPGAEAGALVNASSAAHVNVDTA  
 RA  
 ADLFVSQMMGSR\*  
  
 >gi|529230 UL26 {Herpes simplex virus type 1}  
 35  
  
 {SEQ ID NO:155} = Contig 16 Length: 11707 Type: N Check: 6054  
 ..  
 40 {SEQ ID NO:156} >contig16 (start 190 - stop 2) translated  
 MEAPGIVWVEESVSAITLYAVWLPPRTRDCLHALLYLVCRDAAGEARARFAEVSVGSSXXXXX  
  
 gi|136802|sp|P10192|HEPA\_HSV11 DNA HELICASE/PRIMASE COMPLEX  
 45

[SEQ ID NO:157] >contig16 (start 2855 - stop 240) translated  
 MAETMNVATCTHQTHHAARAPGATSAPGAASGDPLGARRPIGDDECEQYTSSVSLARMLYGGDLAEWV  
 PR  
 VHPKTTIERQQHGPVTFPDASAPTARCVTVVRAPMGSGKTTALIRWLGEAIHSPDTSVLVVSRRSFT  
 5 QT  
 LATRFAESGLPDFVTYFSSTNYIMNDRPFHRLIVQVESLHRVGNLLNNYDVLVLDEVMSLTGLQLYSP  
 TM  
 QQLGRVDALMLRLLRTPRIIAMDATANAQLVDFLCSLRGEKNVHVIGEYAMPGFSARRCLFLPRLG  
 PE  
 10 VLQAALRPPGPAGGAPPPDAPPDATFFGELEARLAGGDNVCIFSSTVSFAEVVARFCRQFTDRVLLH  
 SL  
 TPPGDVTTWGRYRVVIYTTVTVGLSFDPPHFDSMFAYVKPMNYGPDMSVYQSLGRVRTLRKGELLI  
 YM  
 DGSGARSEPVFTPMLLNHVVSASGQWPAQFSQVTNLLCRRFKGRCDASHADAAQARGSRIYSKFRYKH  
 15 YF  
 ERCTLACLADSLNILHMLLTLCNMHVRFWGHDAALTPRNFCLFLRGIHFDALRAQRDLRELRCQDPDT  
 SL  
 SAQAAETEEVGLFVEKYLRPDVAPAEVVALMRGLNSLVGRTRFIYLVLEACLRVPMAAHSSAIFRRL  
 YD  
 20 HYATGVIPTINAAGELELVALHPTLNVAPVWELFRLCSTMAACLQWDSMAGGSGRTFSPEDVLELLNP  
 HY  
 DRYMQLVFELGHCNVTGPLLSEDAVKRVADALSGCPPRGSVSETEHALSLFKIIWGEFGVQLAKST  
 QT  
 FPGAGRVKNLTKRAIVELLDHRIDHSACRTHRQLYALLMAHKREFAGARFKLRAPAWGRCLRTHASG  
 25 AQ  
 PNTDIILEAALSELPTEAWPMMQGAVNFSTL\*

gi|136806|sp|P10193|OBP\_HSV11 ORIGIN OF REPLICATION BINDING

[SEQ ID NO:158] >contig16 (start 2707 - stop 4137) translated  
 VYCSHSSSPMGRRAPRGSPGAAPGADVAPGARAANWVWCVQVATFIVSAICVVGLLVLASVFRDRFPC  
 LY  
 APATSYAEANATVEVRGGVAVPLRLDTQSLLATYAITSTLLAAVYAAVGAVTSRYERALDAARRLA  
 AA  
 35 RMAMPHATLIAGNVCAWLLQITVLLLAHRISQLAHLIYVLHFACLVYLAHFCTRGVLSGTYLQRQVHG  
 LI  
 DPAPTHHRIVGPVRAVMTNALLGTLCTAAAVSLNTIAALNFNFSAPSMCLCLTTLFALLVVSLLL  
 VV  
 EGVLCYVVRVLVGPLGAIAATGIVGLACEHYHTGGYYVVEQQWPGAQTGVRVALALVAAFALAMAVL  
 40 RC  
 TRAYLYHRRHHTKFFVRMRDTRHRAHSALRRVRSSMRGSRRGGPPGDPGYAETPYASVSHAEIDRYG  
 DS  
 DGDPIYDEVAPDHEAELYARVQRPGVPDAEPIYDTVEGYAPRSAGEPVYSTVRRW\*

45 gi|136810|sp|P04288|VGLM\_HSV11 GLYCOPROTEIN

[SEQ ID NO:159] >contig16 (start 4621 - stop 4331) translated  
MGLAFSGARPCCCRHNVIITDGGEVVSLTAHEFDVVDIESEEEGNFYVPPDMRVVTRAPGPQYRRASD  
PP  
5 SRHTRRRDPDVARPPATLTPPLSDSE\*

gi|136816|sp|P13294|UL11\_HSV2 HYPOTHETICAL UL11 PROTEIN

[SEQ ID NO:160] >contig16 (start 6399 - stop 4537) translated  
10 MAAAATPGAKRPADPARDPDSPPKRPRPNSLDLATVFGPRPAPPRPTSPGAPGSHWPQSPPRGQPDGG  
AP  
GEKARPASPALSEASSGPPTPDIPLSPGGAHAIDPDCSPGPPDPDPMWSASAI PNALP PHILAETFER  
HL  
RGLLRGVRSP LAIGPLWARLDYLC SLVVSLEAAGMVDRLGRHLWRLTRAPPSAAEAVAPRPLMGFY  
15 EA  
ATQNQADCQLWALLRRGLTTASTLRWGAQGPCFSSQWLTHNASLRDLAQSSAVMFGRVNEPTARNLLF  
RY  
CVGRADAGVNDDADAGR FVFHQPGDLAEENVHACGVLM DGHTGMVGASLDILVCPRDPHGYLAPAPQT  
PL  
20 AFYEVKCRAKYAFDPADPGAPAASAYEDLMARRSPEAFRAFIRSIPNPGVRYFAPGRVPGPEEALVTQ  
DR  
DWLDSRAAGEKRRCSAPDRALVELNSGVVSEVLLFGVPDLERRTISPAWSSGELVRREPIFANPRHP  
NF  
KQILVQGYVLD SHFPDCPLQPHLVTFLGRHRAGAE EGVTFRLEDGRGAPAGRGGAPGPAKASILPDQA  
25 VP  
IALIITPVRVEPGIYRDIRRNSRLAFDDTLAKLWASRSPGRGPAAADTTSSSPTAGRSSR\*

gi|119694|sp|P06489|EXON\_HSV2 ALKALINE EXONUCLEASE

[SEQ ID NO:161] >contig16 (start 8023 - stop 6440) translated  
30 VGRRRPGGRMDESGRQRPASHVAADIS PQGAHRRSFKAWLAS YIHSLSRRASGRPSGSPRDGAVSGA  
RP  
GSRRRSSFRERLRAGLSRWRVSRSSRRSSPEAPGPAAKLRRPPLRRSETAMTSPSPSPSHILSLARI  
HK  
35 LCIPVFAVN PALRYTTLEIPGARSFGGSGGYGEVQLIREHKLAVKTIREKEWFAVELVATLLVGEAL  
RG  
GRTHDIRGFITPLGFSLQQRQIVFPAYDMDLGKYIGQLASLRATTPSVATALHHCFTDLARAVVFLNT  
RC  
GISHLDIKCANVLM LRS DAVSLRR AVLADFS LVT LNSNSTISRGQFCLQEPDLES PRGFGMPAALT  
40 AN  
FHTLVGHGYNQPPPELLVKYLN NERA EFNNRPLKHDVGLAVDLYALGQTLELLVS VYVAPSLGVPVTR  
VP  
GYQYFNNQLSPDFAVALLAYRCVLHPALFVNSAETNTHGLAYDVPEGIRRHRLNPKIRRAFTEQCINY  
QR  
45 THKAVLSSVSLPPELRPLLVLVSRLCHANPAARHSLS\*

- gi|125628|sp|P04290|KR2\_HSV11 PROBABLE SERINE/THREONINE-PRO
- [SEQ ID NO:162] >contig16 (start 8409 - stop 7750) translated
- 5 MSRDASHAALRRRLAETHLRAEVYRDQTLQLHREGVSTQDPRFVGAFMAAKAAHLELEARLKSRLARLE  
MM  
RQRATCVKIRVEEQAARRDFLTAHRRYLDPALSERLDAADRLADQEEQLEEEAANASLWGDGDLADG  
WM  
SPGDSDDLVMWQLTSAPKVHTDAPSRPGSRPTYTPSAAGRPDAQAAPPPETAPSPEPAPGPAADPASG  
10 SG  
FARDCPDGE\*
- gi|136823|sp|P04291|UL14\_HSV11 HYPOTHETICAL UL14 PROTEIN
- 15 [SEQ ID NO:163] >contig16 (start 8295 - stop 9788) translated  
VYSRPPGVAAGSGPCTPRPGGASRPNVGAGPRGWRLGSSRRPRARPTSDSFAPTPLTSAAPASPAMFG  
QQ  
LASDVQQYLERLEKQRQQKVGVDASAGLTGGDALRVPFLDFATATPKRHQTVVPGVGTLDCCCHS  
PL  
20 FSAVARLLFNLSLVPALRGRDFGGDHTAKLEFLAPELVRAVARLRFRECAPEDAVPQRNAYYSVLNT  
FQ  
ALHRSEAFRQLVHFVRDFAQLLKTSFRASSLAETGPPKKRAKVDVATHGQTYGTLELFQKMILMHAT  
YF  
LAAVLLGDHAEQVNTFLRLVFEIPLFSDTAVRHFRQRATVFLVPRRHGKTWFLVPLIALSLASFRIK  
25 IG  
YTAHIRKATEPVFDEIDACLRGWFSSRVHDHVKGETISFSFPDGSRSTIVFASSHNTNVSTPSSRGAC  
FP  
GAALPEIDRQNTARRECGTTRPQPPPPWRGEALLFCNRTMRLWPRPARPRGSSSLQTGGWYTMTER  
GA  
30 TRRWSSG\*
- gi|139646|sp|P04295|VTER\_HSV11 PROBABLE DNA PACKAGING PROTEIN
- [SEQ ID NO:164] >contig16 (start 10626 - stop 9661) translated
- 35 VWRVVRGDERLKIFRCLTVLTEPLCQVALPDPPERALFCEIFLYLTRPKALRLPSNTFFAIFFFNRE  
RR  
YCATVHLRSVTHPRTPLLCTLAFGHLEAASPPEETPDPAAEQLADEPVAHELDGAYLVPTTEPPNPGA  
CC  
ALGPGAWWHLPGGRIYCWAMDDDLGSLCPPGSRARHLGWLLSRITDPPGGGGACAPTAHIDSANALWR  
40 AP  
AVAEACPCVAPCMWSNMAQRTLAVRGDASLCQLLFGHPVDAVILRQATRRPRITAHLEHVVGVDGAE  
SV  
IRPTSAGWRLCVLSSYTSRLFATSCPAVARAVARASSSDYK\*
- 45 gi|136829|sp|P10200|UL16\_HSV11 PROTEIN UL16

[SEQ ID NO:165] >contig16 (start 11723 - stop 10881) translated  
LTEACAAERVVRPHQLSPAAQTALLRRFPALLEGPLRHPRPVLQPFDAIAEVAFFVARIQIACLRALGHS  
IR  
5 AALQGGPRIFQRLRYDFGPHQSEWLGEVTRRFPVLLLENLMRALEGTAPDAFFHTAYALAVLAHLGGQG  
GR  
GRRRLVPLSDDIPARFADSDAHYAFDYYSTSGDTLRLTNRPVAVVIDGDVNGREQSKCRFMEGSPST  
AP  
10 HRVCEQYLPGESYAYLCGFNRRLCGLVVPFGGFATINTAAYLSLADPVARAVGLRFCRGAATGPGL  
VR

gi|136835|sp|P10201|UL17\_HSV11 PROTEIN UL17

15 [SEQ ID NO:166] = Contig ID 17 Length: 732 Type: N Check: 3911  
..

[SEQ ID NO:167] >contig17 (start 747 - stop 1) translated  
20 PAASPLEPLGDPTLWRALYACVLAALERQTGPVALFVPLRLGWDPTGLVVRVERASWGPPAAPRAAL  
LD  
VEAKVDVDPLALAARVAEHGPARLAWARLAAIRDSPQCASSASLAVTITTRTARFAREYTTLAFPPTS  
KE  
GAFADLVEVCEVGLRPRGHPQRTARVLLPRGYDYFVSAGDGFSAPALVALFRQWHTTVHAAPGALAP  
25 VF  
AFLGAGFDVRGGPVQYFAVLGFPGWPTFTVPAAAXXXXX

30 gi|136802|sp|P10192|HEPA\_HSV11 DNA HELICASE/PRIMASE COMPLEX

[SEQ ID NO:168] = Contig ID 18 Length: 3006 Type: N Check:  
6117 ..

35 [SEQ ID NO:169] >contig18 (start 2 - stop 673) translated  
XXXXXALEREQRAADRAAGGAGRPAAEDLLRADYDIIDVSKSMDDDTYVANSFQHQYIPAYGQDLER  
LS  
40 RLWEHELVRFCILRHRNNQGQETSISYSSGAIASFVAPYFEYVLRAPRAGALITGSDVILGEEELWE  
AV  
FKKTRLQTYLTDVAALFVADVQHAALPRPPSPTPADFRASASPRGGSRSRTRTRSRSRPGRTPRGAPDQ  
GW  
GVERRDGRPHARR\*

45

gi|136794|sp|P10190|UL06\_HSV11 VIRION PROTEIN UL6

[SEQ ID NO:170] >contig18 (start 612 - stop 1538) translated

5 AT VRRTRAGASNAGMADPTPADEGTAAAILKQAIAGDRSLVEVAEGISNQALLRMACEVRQVSDRQPRFT

SVLRVDVTPRGRLRFVLDGSSDDAYVASEDYFKRCGDQPTYRGFAVVVLTANEDHVHSLAVPPLVLLH

RL

SLFRPTDLRDFELVCLLMYLENCPRSHATPSLFVKVSAWLGVARHASPFEVRCLLLRSCHWILNTL

MC

10 MAGVKPFDELVLPHWYMAHYLLANNPPPVLSALFCATPQSSALQLPGVPVPTDCVAYNPAGVMGSCW

KS

KDLRSALVYWWSGSPKRRTSSLFYRFC\*

gi|136798|sp|P10191|UL07\_HSV11 PROTEIN UL7

15 [SEQ ID NO:171] >contig18 (start 3021 - stop 1795) translated

ACLGAWPAVGARVVLPPRAWPAVASEAAGRLLPAFREAVARWHPTATTIQLDPPAAVGPVWTARFCF

SG

LQAQLAALAGLGEAGLPEARFRAGLERLDALVAAAPSEPWAVLERLVPDACDACPALRQLLGGVM

20 AA

VCLQIEQTASSVKFAVCGGTGAFFWGLFNVDPGDADAHGAIQDARRALEASVRAVLSANGIRPLAP

SL

ALEGVYTHVVTWSQTGAFFWNSRDDTDFLQGFPLRGPAYAAAAEVMRDALRRILRRPAAGPPEEAVCA

AR

25 GIMEDACDRFVLDAFGRRLDAEYWSVLTPPGEADDPLPQTAFRGGALLDAEQYWRVVRVCPGGGESV

GV

PVDLYPRPLVLPVDCAHHLREILREIQLVFTGVLEGVWEGGGSFVYPFEKMRFLFP\*

gi|136802|sp|P10192|HEPA\_HSV11 DNA HELICASE/PRIMASE COMPLEX

30

[SEQ ID NO:172] = Contig ID 2 Length: 429 Type: N Check: 5672

..

35 [SEQ ID NO:173] = Contig ID 3 Length: 15901 Type: N Check:

1337 ..

[SEQ ID NO:174] >contig3 (start 1547 - stop 2791) translated

40 MADIPDPALNTTPANHAPSPPPGSRKRRRPVLPSSSESEGKPDTESESSSTESSEDEAGDLRGGR

RR

SPRELGGRYFLDLAESTTGTESEGTGPSDDDDDDASDGWLVDTPPRKSKRPRINLALTSSPDRRAGV

VF

PEVWRNDRPIRAAQPAQSSGDRAAAPRRSARQAQMRSGAAWTLDLHYIRQCVNQLFRILRAAPNP

45 PG



SANRLRHLVRDCYLMGYCRTLGPRTWGRLLQISGGTWDVRLRNAIREVEARFEPAAEPVCELPCLN  
 RR  
 YGPECVGNLETNGGSTSDDEISDATDSDDTLASHSDTEGGPSPAGRENPEASGGAIAARLECEFGT  
 FD  
 5 WTSEEGSQPWLSAVVADTSSAERSGLPAPGACRATEAPEREDGCRKMRFPACPYPCGHTFLRP\*

gi|124184|sp|P04485|IE68\_HSV11 IMMEDIATE-EARLY PROTEIN IE68

[SEQ ID NO:175] >contig3 (start 3848 - stop 2973) translated  
 10 MGVVVVSVVTLTDQRNALPRTSADASPALWSFLLRQCRILASEPLGTPVVVRPANLRRLAEPMLDLPK  
 FT  
 RPIVTRSCRCPPNTTGLFAEDDPLESIIEILDAPACFRLLHQERPGPHRLYHLWVGAADLCVPFLE  
 YA  
 QKTRLGFRFIAMKTNDAWGEPWPLPDRFLPERTVSWTPFPAAPNHPLENLLSRYEYQYGVVVPGDRE  
 15 RS  
 CLRWLRLVAPHNKPRPASSRPHPATHPTQRPCFTCMGRPEIPDEPSWQTGDDDPQNPGPPLAVGDEW  
 PP  
 SSHVCYPITNL\*

20 gi|137125|sp|P13292|US02\_HSV2 PROTEIN US2

[SEQ ID NO:176] >contig 3 (start 4044 - stop 5579) translated  
 VGGCVDKLPLLKTPGPVARGARWLARATRRMACRKFCGVYRRPDKRQEASVPPETNTAPAFPASTFYT  
 PA  
 25 EDAYLAPGPPETIHPSRPPSPGEAARLCQLQEILAQMHSDEEDYPIVDAAGAEEDDEADDDAPDDVAYP  
 ED  
 YAEGRFLSMVSAAPLPGASGHPPVPGRAAPPDVRTCDSGKVGATGFTPEELDTMDREALRAISRCKP  
 PS  
 TLAKLVTGLGFAIHGALIPGSEGCVFDSHPNYPHRVIVKAGWYASTNHEARLLRRLNHPAILPLLDL  
 30 HV  
 VSGVTCLVLPKYHCDLYTYSKRPSPLGHLQITAVSRQLLSAIDYVHCEGIIHRDIKTENILINTPEN  
 IC  
 LGDFGAACFVRGCRSSPFHYGIAGTIDTNAPEVLGDPYTQVIDIWSAGLVIFETAVHTASLFSAPRD  
 PE  
 35 RRPCDNQIARIIRQAQVHVDEFPTHAESRLTAHYRSRAAGNNRPAWTRPAWTRYKYIHTDVEYLICKA  
 LT  
 FDAALRPSAAELLRLPLFHPK\*

40 gi|125617|sp|P13287|KR1\_HSV2 SERINE/THREONINE-PROTEIN KINASE

[SEQ ID NO:177] >contig3 (start 8255 - stop 8368) translated  
 VGGCLMILGMACLLEVLRLRGRELARCCPHAGQFAP\*

45 gi|137132|sp|P13293|VGLJ\_HSV2 GLYCOPROTEIN J

[SEQ ID NO:178] >contig 3 (start 8791 - stop 9993) translated  
 VCIAYHGMGRLTSGVGTAALLVVAVGLRVVCAKYALADPSLKMADPNRFRGKNLPVLDQLTDPGPKR  
 VY  
 HIQPSLEDPFQPPSIPITVYYAVLERACRSVLLHAPSEAPQIVRGASDEARKHTYNLTIAWYRMGDNC  
 5 AI  
 PITVMEYTECPYNKSLGVCPIRTQPRWSYYDSFSAVSEDNLGFLMHAPAFETAGTYLRLVKINDWTEI  
 TQ  
 FILEHRARASCKYALPLRIPPAACLTSKAYQQGVTVDSIGMLPRFIPENQRTVALYSLKIAGWHGPKP  
 PY  
 10 TSTLLPPELSDTTNATQPELVPEDPEDSALLEDPAGTVSSQIPPNWHIPSIQDVAPHHAPAAPSNPGL  
 II  
 GALAGSTLAVLVIGGIAFWVRRRAQMAPKRLRLPHIRDDDAPPSHQPLFY\*

gi|138234|sp|P03172|VGLD\_HSV2 GLYCOPROTEIN D PRECURSOR

15 [SEQ ID NO:179] >contig3 (start 10012 - stop 11313) translated  
 VYLWARVGGWLGYLGGTWTPHKGSLEGGKLGQFIGRERGARTAVPTISHRAHSHLDPSDPMGPGRSLO  
 GL  
 AILGLWVCATGLVVRGPTVSLVSDSLVDAGAVGPQGFVEEDLRVFGELHFVGAQVPHTNYDGIIELF  
 20 HY  
 PLGNHCPRVVHVVTLTACPRRPAVAFTLCRSTHHAHSPAYPTLELGLARQPLLRVRTATRDYAGLYVL  
 RV  
 WVGSA TNASLFLVGLVALSANGTFVYNGSDYGSCDPAQLPFSAPRLGPSSVYTPGASRPTPPRTTSPS  
 SP  
 25 RDPTPAPGDTGTPAPASGERAPPNSTRSASESRHRLTVAQVIQIAIPASIIAFVFLGSCICFIHRCQR  
 RY  
 RRPRGQIYNPGGVSCAVNEAAMARLGAELRSHPNTPPKPRRRSSSTTMPSLTSIAEESSEPGFVLLS  
 VS  
 PRPRSGPTAPQEV\*

30 gi|138328|sp|P06764|VGLI\_HSV23 GLYCOPROTEIN I

[SEQ ID NO:180] >contig3 (start 11632 - stop 12984) translated  
 MARGAGLVFFVGWVWSCLAAPRTSWKRVTSGEDVLLPAPAGPEERTRAHKLLWAAEPLDACGPLR  
 35 PS  
 WVALWPPIRVLETVVDAACMRAPEPLAIAYSPPFPAGDEGLYSELAWRDRVAVVNESLVIYGALETDS  
 GL  
 YTLSVVGLSDEARQVASVVLVVEPAPVPTPTDDYDEEDDAGVSERTFVSVPPPTPPRRPVAPPTH  
 RV  
 40 IPEVSHVRGVTVMETPEAILFAPGETFGTNVSIHAIHDDGPYAMDVVMRFDVPSSCAEMRIYEAC  
 LY  
 HPQLPECLSPADAPCAVSSWAYRLAVRSYAGCSRTTPPPRCFAEARMPEVPGLAWLASTVNLEFQHAS  
 PQ  
 HAGLYLCVVYVDDHIHAWGHMTISTAAQYRNAVVEQHLQORQPEPVEPTRPHVRAPPPAPSARGPLRL  
 45 GA

VLGAALLLAALGLSAWGVHDLAQAALLAGG\*

gi|138240|sp|P04488|VGLE\_HSV11 GLYCOPROTEIN E PRECURSOR

5 [SEQ ID NO:181] >contig3 (start 13431 - stop 13568) translated  
VALHAVDAPSQFVTWLA VRWLRGAVGLGAVLCGIAFYVT SIARGA\*

gi|1944544|gnl|PID|e312381 US8A

10 [SEQ ID NO:182] >contig3 (start 13668 - stop 13937) translated  
MTRSRPADQDSVRSSASVPLYPAA SPVPAEAYYSESEDEAANDFLVRMGRQQSVLRRRRRRTRCVGLVI  
AC  
LVVALLSGGFGALLVWLLR\*

15 gi|135568|sp|P06481|TEGP\_HSV11 TEGUMENT PHOSPHOPROTEIN US9

[SEQ ID NO:183] >contig3 (start 15333 - stop 14425) translated  
MIRRRGNVEIRVYYESVRPSRSRSHLKPSDHQEFPGHHVSPGSPGFPE SPGNREFHDLPENPGSRAYP  
GT  
20 RDPHPHGCPCGSLDPHG NPAQPA GLPSPVPYAPLGSPDPSSPRQRTYVLP RVGIRNAPASDTRAPKRA  
HS  
RHRADRPPE SPGSELYPLNAQALAH LQMLPADHRAFFRTVIEVSRLCALNTHDPPPLAGARVGQEAQ  
LV  
HTQWLRANRESSPLWPWR TAAMNFIAAAAPCVQTHRHMHDLLMACAFWCCLAHASTCSYAGLYSAHCQ  
25 HL  
FRAFGCGPPVLTTSRGQGWCN\*

gi|137138|sp|P06486|US10\_HSV11 VIRION PROTEIN US10

30

[SEQ ID NO:184] = Contig ID 4 Length: 179 Type: N Check: 5124

..

[SEQ ID NO:185] = Contig ID 5 Length: 2117 Type: N Check: 9467

35

..

[SEQ ID NO:186] >contig5 (start 1020 - stop 1) translated  
MLNDMQWLASSDSEETE VGISDDDLHRDSTSEAGSTDTEMFEAGLMDAATPPARPPAERQGSPTPAD  
40 AQ  
GSCGGGPVGEEAEAGGGGDVCAVCTDEIAPPLRCQSFPC LHPFCIPCMKTIWPLRNTCPLCNTPVAY  
LI  
VGVTASGSFSTIPIVNDPRT RVEAEAAVRSGTAVDFIWTGNPRTAPRSLSLGGHTVRALSPTPPWPGT  
DD

EDDDLADGEGGRSGTGRGSGTGRGSGTGRGSGTGRGSGGGRAGVGHWAGVGRGXGTNRGFPSPSA  
AD  
YVPPAPRRAPRRGGGAGATRGTSQPAATRPAPPGAPRSSSSGGAPLRAGVSGSXXXXX

5 gi|124135|sp|P28284|ICP0\_HSV2H TRANS-ACTING TRANSCRIPTIONAL

[SEQ ID NO:187] = Contig 6 Length: 643 Type: N Check: 5042 ..

10

[SEQ ID NO:188] = Contig 7 Length: 354 Type: N Check: 9326 ..

[SEQ ID NO:189] = Contig 8 Length: 6387 Type: N Check: 4794 ..

15

[SEQ ID NO:190] >contig8 (start3 - stop 1454) translated  
XXXXXTRRICARGPALPPGGLAVGGQMYVNRNEIFNAALAVTNIILDLIALKEPVFPRLHEALGHF  
RR  
GALAAVQLLFPAARVDPDAYPCYFFKSACRPRAPPVCAGDGPSAGGDDGDGDWFPDAGGDDGDEEWEE  
20 DT  
DPMDDTHGPLPDDEAAYLDLLHEQIPAATPSEPDSVVCSCADKIGLRVCLPVPAPYVVHGSMTMRGVA  
RV  
IQQAVLLDRDFVEAVGSHVKNFLIDTGVYAHGSLRLPYFAKIGPDGSACGRLLPVFVIPPACEDVP  
AF  
25 VAAHADPRRFHFHAPPMFSAAPREIRVLHSLGGDYVSFFEKKASRNALEHFGRRETLTEVLGRYDVRP  
DA  
GETVEGFASELLGRIVACIEAHFPEHAREYQAVSVRRRAVIKDDWVLLQLIPGRGALNQSLSCLRFKHG  
RA  
SRATARTFLALSVGTNNRLCASLCQQCFATKCDNNRLHTLFTVDAGTPCSRSAPSSTSRPSSS\*

30

gi|136939|sp|P10236|UL52\_HSV11 DNA HELICASE/PRIMASE COMPLEX

[SEQ ID NO:191] >contig8 (start 1406 - stop 2422) translated  
MLAVRSLQHLTTVIFITAYGLVLAWYIVFGASPLHRCIYAVRPAGAHNDTALVWMKINQTLFLGPPT

35

AP  
PGGAWTPHAHVCIYANIEGRAVSLPAIPGAMSRVMNVHEAVNCLEALWDTQMRLVVVGWFLYLAFVA  
LH  
QRRCMFGVVSPAHSMAVAPATYLLNYAGRIVSSVFLQYPYTKITRLLCELSVQRQTLVQLFEADPVTF  
YH  
40 RPAVGVIIVGCELLRFVALGLIVGTALISRGACAITYPFLTITTWCFVSIIALTELYFILRRDSAPK  
NA  
EPAAPRGRSKGWSGVCGRCCSIILSGIAVRLCYIAVVAGVVLMAIRYEQEIQRRLFDL\*

gi|116105|sp|P22485|CELF\_HSV2H CELL FUSION PROTEIN PRECURSOR

45

[SEQ ID NO:192] >contig8 (start 2752 - stop 4506) translated  
VTPDGEGQGGVSESRRSCGYKGSHRPTGRCVLP CADPGCASVPLLSDPATLFRHAPPRRTPAIPAP  
AT  
YNMATDIDMLIDLGLDLS DSELEEDALERDEEGRRDDPESDSSGECSSSDEDMEDPCGDGGAE AIDAA  
5 IP  
KGPPARPEDAGTPEASTPRPAARRGADDPPTTGVWSRLGTRRSASPREPHGGKVARIQPPSTKAPH  
PR  
GRRGRRRRGRGRYGP GGADSTPNPRRRVSRNAHNQGRHPASARTDGP GATHGEARRGGEQLDVSGGP  
RP  
10 RGTRQAPPPLMALSLTPPHADGRAPVPERKAPSADTIDPAVRAVLRSISERAAVERISESFGRSALVM  
QD  
PFGGMPFFAANSPWAPVLATQAGGFDAETRRVSWETLV AHGPSLYRTFAANPRAASTAKAMRDCVLRQ  
EN  
LIEALASADETLAWCKMCIHHNLP LRPQDPIIGTAAAVLENLATRLRPFLQCYLKARGLCGLDDLCSR  
15 RR  
LSDIKDIASFVLVILARLANRVERGVSEIDYTTVGVGAGETMHFYIPGACMAGLIEILDTHRQECSSR  
VC  
ELTASHTIAPLYVHGKYFYCNSLF\*

20 gi|124181|sp|P28276|IE63\_HSV2H TRANSCRIPTIONAL REGULATOR IE

[SEQ ID NO:193] >contig8 (start 4638 - stop 5282) translated  
MWGPGPARFIARPGTHGRRVFTDPPPRNMTTPLSNLFLRAPDITHVAPPYCLNATWQAENALHTTKT  
DP  
25 ACLAARSYLVRASCSTSGPIHCFFFAVYKDSQHSLPLVTELRFADLVNHPPVLRELEDKRGGRRLRCT  
GP  
FSCGTIKDVSGASPAGEYTINGIVYHCHCRYPF SKTCWLGA SAALQHLRSISSSGTAARAAEQRRHKI  
KI  
KIKV\*

30 gi|136947|sp|P28281|UL55\_HSV2H PROTEIN UL55

[SEQ ID NO:194] >contig8 (start 5808 - stop 5455) translated  
MIGAHPGVGGDLPSGLPTYAEATSDRPPTYAMVMAACPTEPPGGSVGPADQPRVQSSRTWRPPLVNSR  
35 EL  
YRAQRAARCASSSDTPQAPGWCGGTCRHAVFGVVAVVVV IILAF LWR\*

gi|136952|sp|P28282|UL56\_HSV2H PROTEIN UL56

40

[SEQ ID NO:195] = Contig 9 Length: 3700 Type: N Check: 8257

45

[SEQ ID NO:196] >contig9 (start 2 - stop 355) translated  
XXXXXGGHAAAGLTELCTLAPRDLTDPLLFAVVGFGVNNHGLMFVVPDIAVYAMLGGAVWISLTQVL  
GL  
RRRLHKDPDAGPWAAATLRGLFFSVYALGFAAGVLVRPRMAASRRSG\*

5 gi|136909|sp|P10227|UL43\_HSV11 MEMBRANE PROTEIN UL43

[SEQ ID NO:197] >contig9 (start 453 - stop 2099) translated  
MGAGVPWTGIKARGAGGPITVRVLGWEVAQKATHPCCSCPREAVVSGNPPRCAGRAHRSFAGAGALLV  
10 MA  
LGRVGLAVGLWGLLVGVVVVLANASPGRTITVGRGNASNAAPSASPRNASAPRTTPTPPQPRKATK  
SK  
ASTAKPAPPKGTGPPKTSSEPVRNCRHDPLARYGSRVQIRCFPNSTRTESRLQIWRYATATDAEIGT  
AP  
15 SLEEVMVNVSAPPGQLVYDSAPNRTDPHVIWAEGAGPGASPRLYSVVGPLGRQRLIIIEELTLETQGM  
YY  
WWWGRTRPSAYGTWVRVRVFRPPSLTIHPHAVLEGQPFKATCTAATYYPGNRAEFVWFEDGRRVDFP  
AQ  
IHTQTQENPDGFSTVSTVTSAAVGGQGPRTFTCQLTWHRDSVSFSRRNASGTASVLPRTITMEFTG  
20 DH  
AVCTAGCVPEGVTFAWFLGDDSSPAEKVAVASQTSCGRPGTATIRSTLPVSYEQTEYICRLAGYPDGI  
PV  
LEHHGSHQPPRPDPTERQVIRAVEGAGIGVAVLVAVVLAGTAVVYLTHASSVRYRRLR\*

25 gi|138220|sp|P06475|VGLC\_HSV23 GLYCOPROTEIN C PRECURSOR

[SEQ ID NO:198] >contig9 (start 2266 - stop 2847) translated  
VGSKRLRKRAPRPDIQARGGAMAFRASGPAYQLAPAAASPARARVPAVAWIGVGAIVGAFALVAALVL  
VP  
30 PRSSWGLSPCDSGWQEFNAGCVAWDPTPVEHEQAVGGCSAPATLIPRAAAKHLAALTRVQAERSSGYW  
WV  
NGDGIRTCLRLVDSVSGIDEFCEELAIRICYPRSPGGFVRFVTSIRNALGLP\*

gi|136917|sp|P06483|UL45\_HSV23 PROTEIN UL45 HOMOLOG

35 [SEQ ID NO:199] >contig9 (start 3716 - stop 3114) translated  
QRPAAARPLAAQREAAGVYDAVRTWGPDAEAEPDQMENTYLLPDDDAAMPAGVGLGATPAADTTAAA  
WP  
AESHAPRAPSEDADSIYESVSEDGGRVYEEIPWVRVYENICLRRQDAGGAAPPGDAPDSPYTEAENPL  
40 YD  
WGSALFSPPGATRAPDPGLSLSPMPARPTNALANDGPTNVAALSALLTKLKRGRHQSH\*

gi|114350|sp|P10230|ATI2\_HSV11 ALPHA TRANS-INDUCING FACTOR

45

TABLE 3

5	[SEQ ID NO:200] = Contig ID 2
10	[SEQ ID NO:201] = Contig ID 3
	[SEQ ID NO:202] = Contig ID 4
15	[SEQ ID NO:203] = Contig ID 5
	[SEQ ID NO:204] = Contig ID 7
20	[SEQ ID NO:205] = Contig ID 12
25	[SEQ ID NO:206] ORF # = 1 from Contig ID 12 ORF start site = 120 ORF end site = 1371 ORF sequence:
30	MADIPDPALNTTPANHAPSPPPGSRKRRRPVLPSSSESEKPDTESESSSTESSEDEAGDLRGGR RR SPRELGGRYFLDLAESTTGTESEGTGPSDDDDDDASDGWLVDTPPRKSKRPRINLRLTSSPDRRAGV VF PEVWRNDRPIRAAQPAQAQSSGDRAAAPRRSARQAQMRSGAAWTLDLHYIRQCVNQLFRILRAAPNP
35	PG SANRLRHLVRDCYLMGYCRTLGPRTWGRLLQISGGTWDVRLRNAIREVEARFEPAAEPVCELPCLNA RR YGPECDVGNLETNGGSTSDDEISDATSDDTLASHSDTEGGPSPAGRENPEASGGATAARLECEFGT FD
40	WTSEEGSQPWL SAVVADTSSAERSGLPAPGACRATEAPEREDGCRKMRFFAACPYPCGHTFLRP*
	Gene matched:gi 124184 sp P04485 IE68_HSV11 Gene name: IMMEDIATE-EARLY PROTEIN IE68
45	[SEQ ID NO:207]

ORF # = 2 from Contig 12

ORF start site = 2428

ORF end site = 1553

ORF sequence:

5 MGVVVSVVTLDDQRNALPRTSADASPALWSFLLRQCRILASEPLGTPVVVRPANLRLAEPLMDLPK  
FT  
RPIVRTRSCRCPPNTTTGLFAEDDPLEISIEILDAPACFRLHQRPGPHRLYHLWVVGAAADLCVPFLE  
YA  
QKTRLGFRFIAMKTNDWVGEPLPDRFLPERTVSWTPFPAAPNHLENLLSRYEYQYGVVVPGDRE  
10 RS  
CLRWLRLSLVAPHNKPRPASSRPHPATHPTQRPCFTCMGRPEIPDEPSWQTGDDDPQNP GPPLAVGDEW  
PP  
SSHVCYPITNL\*

15 Gene matched: gi|137125|sp|P13292|US02\_HSV2

Gene name: PROTEIN US2.gi|419137|pir|A4

[SEQ ID NO:208]

ORF # 3 from Contig 12

20 ORF start site = 2714

ORF end site = 4159

ORF sequence:

MACRKFCGVYRRPDKRQEASVPPETNTAPAFPASTFYTPAEDAYLAPGPPETIHPSRPPSPGEAARLC  
QL  
25 QEILAQMHSDDEDYPIVDAAGAEEDDEADDDAPDDVAYPEDYAEGRFLSMVSAAPLPGASGHPPVPGRA  
AP  
PDVRTCDSGKVGATGFTPEELDTMDREALRAISRCKPPSTLAKLVTGLGFAIHGALIPGSEGCVFDS  
SH  
PNYPHRVIVKAGWYASTNHEARLLRRLNHPAILPLDLHVVSQVTCVLPLKYHCDLYTYSKRPSPLG  
30 HL  
QITAVSRQLLSAIDYVHCEGIIHRDIKTENILINTPENICLGDFGAACFVRGCRSSPFHYGIAGTIDT  
NA  
PEVLADPYTQVIDIWSAGLVIFETAVHTASLFSAPRDPERRPCDNQIARIIRQAQVHVDEFPTHAES  
RL  
35 TAHYRSRAAGNNRPAWTRPAWTRYKIHTDVEYLICKALTFDAALRPSAAELLRLPLFHPK\*

Gene matched: gi|125617|sp|P13287|KR1\_HSV2

Gene name: SERINE/THREONINE-PROTEIN KINAS

40 [SEQ ID NO:209]

ORF # = 4 from Contig 12

ORF start site = 6835

ORF end site = 6948

ORF sequence:

45 VGGLCLMILGMACLLEVLRLRGRELARCCPHAGQFAP\*



Gene matched: gi|137132|sp|P13293|VGLJ\_HSV2

Gene name: GLYCOPROTEIN J.gi|419140|pir|

5

{SEQ ID NO:210}

ORF # = 5 from Contig 12

ORF start site = 7392

10 ORF end site = 8573

ORF sequence:

MGRLTSGVGTAALLVVAVGLRVVCAKYALADPSLKMADPNRFRGKNLPVLDQLTDPGKRVYHIQPS  
LE

15 DPFQPPSIPITVYAVLERACRSVLLHAPSEAPQIVRGASDEARKHTYNLTIAWYRMGDNCAIPITVM

EY  
TECPYNKSLGVCPIRTQPRWSYYSFSAVSEDNLGFLMHAPAFETAGTYLRLVKINDWTEITQFILEH  
RA

RASCKYALPLRIPPAACLTSKAYQQGVTVDSIGMLPRFIPENQRTVALYSLKIAGWHGPKPPYTSTLL  
PP

20 ELSDTTNATQPELVPEDPEDSALLEDPAGTVSSQIPPNWHIPSIQDVAPHHAPAAPSNPGLIIGALAG  
ST

LAVLVIGGIAFWVRRRAQMAPKRLRLPHIRDDDAPPSHQPLFY\*

Gene matched: gi|138234|sp|P03172|VGLD\_HSV2

25 Gene name: GLYCOPROTEIN D PRECURSOR

{SEQ ID NO:211}

ORF # = 6 from Contig 12

ORF start site = 8775

30 ORF end site = 9893

ORF sequence:

MPGRSLQGLAILGLWVCATGLVVRGPTVSLVSDSLVDAGAVGPQGFVEEDLRVFGELHFVGAQVPHTN  
YY

35 DGIIELFHYPLGNHCPRVVHVVTLTACPRRPAVAFTLCRSTHHAHSPAYPTLELGLARQPLLVRTAT  
RD

YAGLYVLRVWVGSATNASLFLGVALSANGTFVYNGSDYGSCDPAQLPFSAPRLGPSSVYTPGASRPT  
PP

RTTTSPPSSPRDPTPAPGDTGTAPASGERAPPNSTRSASESRHRLTVAQVIQIAIPASIIAFVFLGSC  
IC

40 FIHRCQRRYRRPRGQIYNPGGVSCAVNEAAMARLGAELRSHPNTPPKPRRRSSSSTTMPSLTSIAEES  
EP

GPVVLLSVSPRPRSGPTAPQEV\*

Gene matched: gi|138328|sp|P06764|VGLI\_HSV23

45 Gene name: GLYCOPROTEIN I^Agi|73722|pir|

[SEQ ID NO:212]  
 ORF # = 7 from Contig 12  
 5 ORF start site = 10212  
 ORF end site = 11858  
 ORF sequence:  
 MARGAGLVFFVGWVVSCLAAAPRTSWKRVTSGEDVLLPAPAGPEERTRAHKLLWAAEPLDACGPLR  
 PS  
 10 WVALWPPIRVLETVVDAACMRAPEPLAIAYSPPFPAGDEGLYSELAWRDRVAVVNESLVIYGALETDS  
 GL  
 YTLSSVGLSDEARQVASVVLVVEPAPVPTPTPDDYDEEDDAGVSERTPVSVPPPTPPRGPPVAPPTH  
 RV  
 IPEVSHVRGVTVMETPEAILFAPGETFGTNVSIHAIHDDGPYAMDVVMRFDVPSSCAEMRIYEAC  
 15 LY  
 HPQLPECLSPADAPCAVSSWAYRLAVRSYAGCSRTTPPPRCFAEARMPEVPGLAWLASTVNLEFQHAS  
 PQ  
 HAGLYLCVVYVDDHIHAWGHMTISTAAQYRNAVVEQHLPPRQPEPVEPTRPHVRAPPPAPSARGPLRL  
 GA  
 20 VLGAALLLAALGLSAWACMTCWRRRSWRAVKSRSATGPTYIRVADSELYADWSSDSEGERDGLWQD  
 PP  
 ERPDSPSTNGSGFEILSPTAPSVYPHSEGRKSRRPLTTFGSGSPGRRHSQASYSSVLW\*

25 Gene matched: gi|138240|sp|P04488|VGLE\_HSV11  
 Gene name: GLYCOPROTEIN E PRECURSOR

[SEQ ID NO:213]  
 ORF # = 8 from Contig 12  
 30 ORF start site = 12010  
 ORF end site = 12147  
 ORF sequence:  
 VALHAVDAPSQFVTWLAVRWLRGAVGLGAVLCGIAFYVTSIARGA\*

35 Gene matched: gi|1944544|gnl|PID|e312381  
 Gene name: (X14112) US8A [human herpesvirus]

40 [SEQ ID NO:214]  
 ORF # = 9 from Contig 12  
 ORF start site = 12247  
 ORF end site = 12516  
 ORF sequence:

MTSRPADQDSVRSSASVPLYPASPVPAAEAYYSESEDEAANDFLVRMGRQQSVLRRRRRTRCVGLVI  
AC  
LVVALLSGGFGALLVWLLR\*

5

Gene matched: gi|135568|sp|P06481|TEGP\_HSV11  
Gene name: TEGUMENT PHOSPHOPROTEIN US9

10

[SEQ ID NO:215]  
ORF # = 10 from Contig 12  
ORF start site = 13912  
ORF end site = 13004  
ORF sequence:

15

MIRRRGNVEIRVYVESVRPSRSRSHLKPSDHQEFPGHHVSPGSPGFPESPGNREFHDLPENPGSRAYP  
GT  
RDPHDPHGCPGSLDPHGNAQAGLPSPVPYAPLGSPDPSSPRQRTYVLPVVGIRNAPASDTRAPKRA  
HS

20

RHRADRPPESPGSELYPLNAQALAHQLMLPADHRAFFRTVIEVSRLCALNTHDPPPPLAGARVGQEAQ  
LV  
HTQWLRANRESSPLWPWRTAAMNFIAAAAPCVQTHRHMHDLLMACAFWCCLAHASTCSYAGLYSAHCQ  
HL  
FRAFGCGPPVLTTSRGQGGWCN\*

25

Gene matched: gi|137138|sp|P06486|US10\_HSV11  
Gene name: VIRION PROTEIN US10

30

[SEQ ID NO:216]  
ORF # = 11 from Contig 12  
ORF start site = 15899  
ORF end site = 16582  
ORF sequence:

35

MSAEQRKKKKTTTTTQGRGAEVAMADEDGGRRLAAAETTGGPGSPDPADGPPPTPNPDRRPAARPGFG  
WH  
GGPEENEDEDDAAADADADEAAPASGEAVDEPAADGVVSPRQLALLASMVDEAVRTIPSPPPERDGA  
EE  
EAARSPSPRTPSMCADYGEENDDDDDDDRDAGRWRGPENDVRGPRGVPGPHGQPVAATPGAPPTP

40

PP  
PPPPPPPARPPPALDRL\*

Gene matched: gi|124141|sp|P08392|ICP4\_HSV11  
45 Gene name: TRANS-ACTING TRANSCRIPTIONAL

[SEQ ID NO:217] = Contig ID 15

5

[SEQ ID NO:218]

ORF # = 1 from Contig 15

ORF start site = 755

ORF end site = 1297

10 ORF sequence:

MRTPADDVSWRYEAPSVIDYARIDGIFLRYHCPGLDTFLWDRHAQRAYLVNPFLFAAGFLEDLSHSVF  
PA  
DTQETTTRRALYKEIRDALGSRKQAVSHAPVRAGCVNFDYSRTRRCVGRDLRPANTTSTWEPPVSSD  
DE

15 ASSQSKPLATQPPVLALSNAPPRRVSPTRGRRRHTRLRRN\*

Gene matched: gi|136776|sp|P28278|VGLL\_HSV2H

Gene name: GLYCOPROTEIN L PRECURSOR•gi|

20

[SEQ ID NO:219]

ORF # = 2 from contig 15

ORF start site = 1170

25 ORF end site = 2174

ORF sequence:

MKRARSRSPPSPSRPSSPFRTPPHGGSPRREVAGILASDATSHVCIASHPGSGAGYPTRLAAGSAVQ  
RR  
RPRGCPPGVMFSASTTPEQPLGLSGDATPPLPTSVPLDWAAFRRRAFLIDDAWRPLLEPELANPLTARL  
30 LA  
EYDRRCQTEEVLPREDVFSWTRYCTPDDVRVVIIGQDPYHHPGQAHGLAFSVRADVPVPPSLRNVLA  
AV  
KNCYPDARMSGRGCLEKWARDGVLLNLTTLTVKRGAAASHSKLGWDRFVGGVVRRLAARRPGLVFMLW  
GA

35 HAQNAIRPDPRQHYYVLKFSHPSPLSKVPFGTCQHFLAANRYLETRDIMPIDWSV\*

Gene matched: gi|137037|sp|P10186|UNG\_HSV11

Gene name: URACIL-DNA GLYCOSYLASE

40

[SEQ ID NO:220]

ORF # = 3 from Contig 15

ORF start site = 2229

45 ORF end site = 2930

ORF sequence:

MVKSRVSYRSVMSGVEERVPSAFTILASWGWFAPQNHDPGASPNTTPIESIAGTAPDAHVGPLDGE  
PD  
RDAISPLTSSVAGDPPGADGPYVTFDTLFMVSSIDELGRRQLTDTIRKDLRLSLAKFSIACHTKTSSFS  
5 GT  
AARQRKRGAPPQRTCVPKSNKSLQMFVLCKRANAAQVREQLRAVIRSRKPRKYYTRSSDGRLCPAVPV  
FV  
HEFVSSEPMRLHRDNLVMLSTEPD\*

10

Gene matched: gi|136782|sp|P28279|UL03\_HSV2H

Gene name: PROTEIN UL3

15

[SEQ ID NO:221]

ORF # = 4 from contig 15

ORF start site = 3735

ORF end site = 3130

ORF sequence:

20

MGNPQTTIAYSLHHPRASLTALPDAAQVVHVFESGTRAVLTRGRARQDRLPRGGVVIQHTPIGLLVI  
ID  
CRAEFCAIRFIGRASTQRLERWWDAMHAYPFDSWSSSHGESVRSATAGILTVVWTPDTIYITATY  
GT  
APEAARGCDNAPLDVRPTTPAPVSPTAGEFPANTTDLLVEVLREIQISPTLDDADPTPGT\*

25

Gene matched: gi|136788|sp|P28280|UL04\_HSV2H

Gene name: PROTEIN UL4.gi|73890|pir|WM

30

[SEQ ID NO:222]

ORF # = 5 from Contig 15

ORF start site = 6447

ORF end site = 3802

35

ORF sequence:

MAASGGEGSRDVRAPGPPPPQPGARPAVFRDEAFNLFTSMHGVQPIIARIRELSQQQLDVTQVPRLO  
WF  
RDVAALEVPTGLPLREFPPFAAYLITGNAGSGKSTCVQTLNEVLDCVVTGATRIAAQNMYVKLSGAFLS  
RP  
40 INTIFHEFGFRGNHVQAQLGQHPYTLASSPASLEDLQRRDLTYWVILDTIKRALAAHGGEARNEF  
HA  
LTALEQTLGLGQALTRLASVTHGALPAFTRSNIIVIDEAGLLGRHLLTTVVYCWWMINALYHTPQYA  
GR  
LRPVLVCVGSPTQTASLESTFEHQKLRCQSVRQSENVLTYLICNRTLREYTRLSHSWAIFINNKRCEVH  
45 EF

GNLMKVLEYGLPITEEHMQFVDRFVVPESYITNPANLPGWTRLFSSHKEVSAYMAKLHAYLKVTRERE  
 FV  
 VFTLPVLTFSVKEFDEYRRLTQQPTLTMEKWITANASRITNYSQSQDQDAGHVRCEVHKSQQLVVAR  
 ND  
 5 ITYVLNSQVAVTARLRKMVFGDGTFRTFEAVLRDDSFVKQTQGETSVEFAYRFLSRLMFGGLIHFYNF  
 LQ  
 RPGLDATQRTLAYGRIGELTAELLSLRRDAAGASATRAADTSDRSPGERAFNFKHLGPRDGGPDDFPD  
 DD  
 LDVIFAGLDEQQLDVFYCHYALEEPETTAAVHAQFGLLKRAFLGRYLILRELFGEVFESAPFSTYVDN  
 10 VI  
 FRGCELLTGSPRGGLMSVALQTDNYTLMGYTYTRVFATAEELRRRHATAGVAEFLEESPLPYIVLRDQ  
 HG  
 FMSVVNTNISEFVESIDSTELAMAINADYGISSKLAMTITRSQGLSLDKVAICFTPGNLRNLNSAYVAM  
 SR  
 15 TTSSEFLHMNLNPLRERHERDDVISEHILSALRDPNVVIVY\*

Gene matched: gi|122809|sp|P10189|HELI\_HSV11

Gene name: PROBABLE HELICASE

20

{SEQ ID NO:223}

ORF # = 7 from Contig 15

ORF start site = 8457

25 ORF end site = 9347

ORF sequence:

MADPTPADEGTAAAILKQAIAGDRSLVEVAEGISNQALLRMACEVRQVSDRQPRFTATSVLRVDVTPR  
 GR

LRFVLGSSDDAYVASEDYFKRCGDQPTYRGFAVVVLTANEDHVHSLAVPPLVLLHRLSLFRPTDLRD

30 FE

LVCLLMYLENCPRSHATPSLFBKVSALGVVARHASPFEVRCLLLRSCHWILNTLMCMAGVKPFDDDE  
 LV

LPHWYMAHYLLANNPPVLSALFCATPQSSALQLPGPVPRPTDCVAYNPAGVMGSCWKSADLRSALVYW  
 WL

35 SGSPKRRTSSLFYRFC\*

Gene matched: gi|136798|sp|P10191|UL07\_HSV11

Gene name: PROTEIN UL7

40 {SEQ ID NO:224}

ORF # = 8 from Contig 15

ORF start site = 11855

ORF end site = 9604

ORF sequence:

MEAPGIVWVEESVSAITLYAVWLPPRTRDCLHALLYLVCRDAAGEARARFAEVSVGSSDLQDFYGSPD  
VS  
AAGAVAAARAAPASPLEPLGDPTLWRALYACVLAALERQTGPVALFVPLRLGWDPQTGLVVRVERAS  
WG  
5 PPAAPRAALLDVEAKVDVDPLALAAARVAEHGPARLAWARLAAIRDSPQCASSASLAVTITTRTARFAR  
EY  
TTLAFPPPTSKEGAFADLVEVCEVGLRPRGHPQRVTARVLLPRGYDYFVSAGDGFSA PALVALFRQWHT  
TV  
HAAPGALAPVFAFLGPGFEVRGGPVQYFAVLGFPGWPTFTVPAAAAAESARDLVRGAAATHAACLGAW  
10 PA  
VGARVVLPPRAWPAVASEAAGRLLPAFREAVARWHPTATTIQLLDPPAAVGPVWTARFCFSGLQAQLL  
AA  
LAGLGEAGLPEARGRAGLERLDALVAAAPSEPWAVLERLVPDACDACPALRQLLGGVMAAVCLQIE  
QT  
15 ASSVKFAVCGGTGAAFWGLFNVDPGDADAAHGAIQDARRALEASVRAVLSANGIRPLAPSLALEGVY  
TH  
VVTWSQTGAWFNSRDDTDFLQGFPLRGPAYAAAAEVMRDALRRILRRPAAGPPEEAVCAARGIMEDA  
CD  
RFVLDAFGRRDLAEYWSVLTPGEADDPLPQTAFRGALLDAEQYWRRVVRVCPGGGESVGVVPDLYP  
20 RP  
LVLPPVDCAHHLREILREIQLVFTGVLEGVWEGGGSFVYPFECKMRFLFP\*

Gene matched: gi|136802|sp|P10192|HEPA\_HSV11

Gene name: DNA HELICASE/PRIMASE COMPLEX

25

[SEQ ID NO:225]

ORF # = 10 from Contig 15

ORF start site = 14399

30 ORF end site = 15802

ORF sequence:

MGRRAPRGSPAAAPGADVAPGARAAWVWCVQVATFIVSAICVVGLLVLASVFRDRFPCLYAPATSYA  
EA  
NATVEVRGGVAVPLRLDTQSLLATYAITSTLLAAAVYAAVGAVTSRYERALDAARRLAAARMAMPHA  
35 TL  
IAGNVCAWLLQITVLLLAHRISQLAHLIYVLHFACLVYLAHFCTRGVLSGTYLRQVHGLIDPAPTHH  
RI  
VGPVRAVMTNALLGTLCTAAAVSLNTIAALNFNFSAPSMILCTTLFALLVVSLLLVEGVLCYH  
VR  
40 VLVGPHLGAIATGIVGLACEHYHTGGYYVVEQQWPGAQTGVRVALALVAAFALAMAVLRCTRAYLYH  
RR  
HHTKFFVRMRDTRHRAHSALRRVRSSMRGSRRGPPGDPGYAETPYASVSHHAEIDRYGDSGDPIYD  
EV  
APDHEAELYARVQRPGVPDAEPIYDTVEGYAPRSAGEPVYSTVRRW\*

45

Gene matched: gi|136810|sp|P04288|VGLM\_HSV11

Gene name: GLYCOPROTEIN M

5 [SEQ ID NO:226]

ORF # = 11 from Contig 15

ORF start site = 16286

ORF end site = 15996

ORF sequence:

10 MGLAFSGARPCCRHNVIIITDGGEVVSLETAHEFDVVDIESEEEGNFYVPPDMRVVTRAPGPQYRRASD  
PP  
SRHTRRRDPDVARPPATLTPPLSDSE\*

15 Gene matched: gi|136816|sp|P13294|UL11\_HSV2

Gene name: HYPOTHETICAL UL11 PROTEIN

[SEQ ID NO:227]

20 ORF # = 12 from Contig 15

ORF start site = 18064

ORF end site = 16202

ORF sequence:

MAAAATPGAKRPADPARDPDSPPKRPRPNSLDLATVFGPRPAPPRPTSPGAPGSHWPQSPPRGQPDGG  
25 AP  
GEKARPASPALSEASSGPPTPDIPSPGGAHAIDPDCSPGPPDPDPMWSASAIPNALPPHILAETFER  
HL  
RGLLRGVRSPLAIGPLWARLDYLC SLVVSLEAAGMVDRLGRHLWRLTRRAPPSAAEAVAPRPLMGFY  
EA  
30 ATQNQADCQLWALLRRGLTTASTLRWGAQGPCFSSQWLTHNASLRLDAQSSAVMFGRVNEPTARNLLF  
RY  
CVGRADAGVNDDADAGR FVFHQPGDLAEENVHACGVLMDGHTGMVGASLDILVCPRDPHGYLAPAPQT  
PL  
AFYEVKCRAKYAFDPADPGAPAASAYEDLMARRSPEAFRAFIRSIPNPGVRYFAPGRVPGPEEALVTQ  
35 DR  
DWLDSRAAGEKRRC SAPDRALVELNSGVVSEVLLFGVPDLERRTISPVAWSSGELVRREPIFANPRHP  
NF  
KQILVQGYVLDSHFDCPLQPHLVTFLGRHRAGAE EGVTFRLEDGRGAPAGRGGAPGPAKASILPDQA  
VP  
40 IALIIITPVRVEPGIYRDIRRNSRLAFDDTLAKLWASRSPGRGPAAADTTSSSPTAGRSSR\*

Gene matched: gi|119694|sp|P06489|EXON\_HSV2

Gene name: ALKALINE EXONUCLEASE•gi|33025

45



[SEQ ID NO:228]  
ORF # = 13 from Contig 15  
ORF start site = 19661  
5 ORF end site = 18107  
ORF sequence:  
MDESGRQRPASHVAADISPOGAHRRSFKAWLASYIHSLSRRASGRPSGSPRDGAVSGARPGSRRRSS  
FR  
ERLRAGLSRWVSRSSRRRSSPEAPGPAKLRRPPLRRSETAMTSPSPSPSHILSLARIHKLCIPVFA  
10 VN  
PALRYTTLEIPGARSFGGSGGYGEVQLIREHKLAVKTIREKEWFAVELVATLLVGECALRGGRTHDIR  
GF  
ITPLGFSLQQRQIVFPAYDMDLGKYIGQLASLRATTPSVATALHHCFTDLARAVVFLNTRCGISHLDI  
KC  
15 ANVLVMLRSDAVSLRAVLADFSVLTLSNSTISRGQFCLQEPDLESPRGFGMPAALT TANFHTLVGH  
GY  
NQPPPELLVKYLNNERAEFNNRPLKHDVGLAVDLYALGQTLLELLVSVYVAPSLGVPVTRVPGYQYFNN  
QL  
SPDFAVALLAYRCVLHPALFVNSAETNTHGLAYDVPEGIRRHLPKIRRAFTEQCIN YQRTHKAVLS  
20 SV  
SLPPELRPLLVLVSRLCHANPAARHSLS\*

Gene matched: gi|125628|sp|P04290|KR2\_HSV11  
Gene name: PROBABLE SERINE/THREONINE-PROTEIN KINASE  
25

[SEQ ID NO:229]  
ORF # = 14 from Contig 15  
ORF start site = 20074  
30 ORF end site = 19415  
ORF sequence:  
MSRDASHAALRRRLAETHLRAEVYRDQTLQLHREGVSTQDPRFVGAFMAAKAAHLEEARLKSRARLE  
MM  
RQRATCVKIRVEEQAARRDFLTAHRRYLDPALSERLDAADDR LADQEEQLEEAANASLWGDGDLADG  
35 WM  
SPGDSDLLVMWQLTSAPKVHTDAPSRPGSRPTYTPSAAGRPDAQAAPPPETAPSPEPAPGPAADPASG  
SG  
FARDCPDGE\*

40  
Gene matched: gi|136823|sp|P04291|UL14\_HSV11  
Gene name: HYPOTHETICAL UL14 PROTEIN

45 [SEQ ID NO:230]

ORF # = 15 from Contig 15

ORF start site = 20155

ORF end site = 21453

ORF sequence:

5 MFGQQ  
 LASDVQQYLERLEKQRQQKVGVEASAGLTGGDALRVPFLDFATATPKRHQTVVPGVGTLDHCCCHS  
 PL  
 FSAVARLLFNSLVPAQLRGRDFGGDHTAKLEFLAPELVRAVARLRFRECAPEDAVPQRNAYYSVLNT  
 FQ  
 10 ALHRSEAFRQLVHFVRDFAQLLKTSFRASSLAETTGPCKRAKVDVATHGQTYGTLELFQKMILMHAT  
 YF  
 LAAVLLGDHAEQVNTFLRLVFEIPLFSDTAVRHFRQRTVFLVPRRHGKTWFLVPLIALSLASFRGIK  
 IG  
 YTAHIRKATEPVFDEIDACLRGWFSSRVDHVKGETISFSFPDGSRTIVFASSHNTNVSTPSSRGAC  
 15 FP  
 GAALPEIDRQTNTARRECGTTRPQPPPPWRGEALLFICNRTMRLWPRPARPRGSSLQTGGWYTMTER  
 GA  
 TRRWSSG\*

20 Gene matched: gi|139646|sp|P04295|VTER\_HSV11  
 Gene name: PROBABLE DNA PACKAGING PROTEIN

[SEQ ID NO:231]

25 ORF # = 16 from Contig 15  
 ORF start site = 22291  
 ORF end site = 21326  
 ORF sequence:  
 VWRVVRGDERLKIFRCLTVLTEPLCQVALPDPDPERALFCEIFLYLTRPKALRLPSNTFFAIFFFNRE  
 30 RR  
 YCATVHLRSVTHPRTPLLCTLAFGHLEAASPPEETPDPAEQLADEPVAHELDGAYLVPTPEPPNPGA  
 CC  
 ALGPGAWWHLPGGRIYCWAMDDDLGSLCPPGSRARHLGWLLSRITDPPGGGGACAPTAHIDSANALWR  
 AP  
 35 AVAEACPCVAPCMWSNMAQRTLAVRGDASLCQLLFGHPVDAVILRQATRRPRITAHLEVVVGRDGAE  
 SV  
 IRPTSAGWRLCVLSSYTSRLFATSCPAVARAVARASSSDYK\*

40 Gene matched: gi|136829|sp|P10200|UL16\_HSV11  
 Gene name: PROTEIN UL16

[SEQ ID NO:232]

45 ORF # = 17 from Contig 15

ORF start site = 24654  
ORF end site = 22546  
ORF sequence:  
MNAHFANEVQYDLTRDPSSPASLIHVIIISSECLAAAGVPLSALVRGRPDGGAAANFRVETQTRAHATG  
5 DC  
TPWRSFAAAYVPADAVGAILAPVIPAHPDLLPRVPSAGGLFVSLPVACDAQGVDPYTVAALRLAWGP  
WA  
TCARVLLFSYDELVPPNTRYAADGARLMRLCRHFCRYVARLGAAAPAAATEAAAHLSLGMGESGTPTP  
QA  
10 SSVSGGAGPAVVGTPDPPIISPEEQLTAPGGDTATAEDVSITQENEEILALVQRAVQDVTRRHFPVRARP  
KH  
AASGVASGLRQALVHQAVSGGALGASDAEAVLAGLEPPGGGRFATPGGPRAAGEDVLNDVLTLPVGT  
AK  
PRSLVEWLDRGWEALAGGDRPDWLWSRRSISVVLRRHHYGTKQRFVVVSYENSVAWGGRRARPPLSSE  
15 LA  
TALTEACAAERVVRPHQLSPAAQTALLRRFPALLEGPLRHPRPVLQPFDAAEVAFVARIQIACLRALG  
HS  
IRAAALQGGPRIFQRLRYDFGPHQSEWLGEVTRRFVLLLENLMRALEGTAPDAFFHTAYALAVLAHLGG  
QG  
20 GRGRRRLVPLSDDIPARFADSDAHYAFDYYSTSGDTLRLTNRPVIAVVIDGDVNGREQSKCRFMEGSP  
ST  
APHRVCEQYLPGESYAYLCLGFNRRLCGLVVPFGGFAPTINTAAYLSLADPVARAVGLRFCRGAATGP  
GL  
VR\*  
25

Gene matched: gi|136835|sp|P10201|UL17\_HSV11  
Gene name: PROTEIN UL17

30 [SEQ ID NO:233]  
ORF # = 18 from Contig 15  
ORF start site = 24684  
ORF end site = 25955  
35 ORF sequence:  
VPEGAWVGGACARPRGPRAHVRLYAVCFVCPQGIRGQDFNLLFVDEANFIRPDAVQTIMGFLNQANCK  
II  
FVSSTNTGKASTSFLYNLRGADELLNVVTYICDDHMPRVVTHTNATACSCYILNKPVFITMDGAVRR  
TA  
40 DLFLPDSFMQEIIIGGQARETGDDRPVLTKSAGERFLLYRPSTTTNSGLMAPELYVYVDPAPTANTRAS  
GT  
GIAVVGRYRDDFIIFALEHFFLRALTGSAPADIARCVVHSLAQVLALHPGAFRSVRVAVEGNSSQDSA  
VA  
IATHVHTEMHRILASAGANGPGPELLFYHCEPPGAVLYPFFLLNKQKTPAFEYFIKKFNSSGGVMASQ  
45 EL

VSVTVRLQTD PVEYLSEQLNNLIETVSPNTDVRMYSGKRNGAADDLMVAVIMAIYLAAPTGIPPAFFP  
IT  
RTS\*

5

Gene matched: gi|139646|sp|P04295|VTER\_HSV11  
Gene name: PROBABLE DNA PACKAGING PROTEIN

[SEQ ID NO:234]

10

ORF # = 19 from Contig 15

ORF start site = 27251

ORF end site = 26295

ORF sequence:

MITDCFEADIAIPSGISRPDAAALQRCEGRVFLPTIRRQLALADVAHESFVSGGVSPDTLGLLLAYR

15

RR

FPAVITRVLPTTRIVACPVDLGLTHAGTVNLRNTSPVDLCNGDPVSLVPPVFEGQATDVRLES�DLTLR  
FP

VPLPTPLAREIVARLVARGIRDLPDPRTPGELPDLNVLYNGARLSLVADVQQLASVNTELRSLVLN  
MV

20

YSITEGTTLILTLPRL LALS AQDGYVNALLQM QSVTREAAQLIHPEAPMLMQDGERRLP LYEALVAW  
LA

HAGQLGDILALAPAVRVCTFDGAAVVQSGD MAPVIRYP\*

Gene matched: gi|139191|sp|P10202|VP23\_HSV11

25

Gene name: CAPSID PROTEIN VP23

[SEQ ID NO:235]

ORF # = 21 from Contig 15

30

ORF start site = 32735

ORF end site = 32067

ORF sequence:

MTMRDDVPLLDREL VYEAACGGEDGELPLDEQFSLSSYGTSDFFVSSAYSRLPPHTQPVFSKRVMFA  
WS

35

FLVLKPLELVAAGMYYGWTGRAVAPACIIAAVLAYYVTW LARALLLVN IKRDR LPLSPPVFWGLCVI  
MG

GAALCALVAAAHETFS PDGLFHWITASQLLPRTDPLRARS LGIACAAGAMWVAAADCFAAFTNFFLA  
RF

WTRAILKAPVAF\*

40

Gene matched: |136841|sp|P10204|UL20\_HSV11

Gene name: MEMBRANE PROTEIN UL20

45

[SEQ ID NO:236]

ORF # = 23 from Contig 15

ORF start site = 37721

ORF end site = 35205

5 ORF sequence:

MGPGLWVVMGVLVGAGGHDTYWTEQIDPWFLHGLGLARTYWRDTNTGRLWLPNTPDASDPQGRGLAP  
PG

ELNLTASVPMRLRYAERFCFVLVTTAEFPRDPGQLLYIPKTYLLGRPRNASLPELPEAGPTSRRPAE  
VT

10 QLKGLSHNPGASALLRSRAWVTFAAAPDREGLTFPRGDDGATERHPDGRRNAPPPGPPAGTPRHPTTN  
LS

IAHLHNASVTWLAARGLLRTPGRYVYLSPSASTWPVGWTTGGLAFGCDAALVRARYGKGFMGLVISM  
RD

15 SPPAEIIVVPADKTLARVGNPTDENAPAVLPGPPAGPRYRVFVLGAPTPADNGSALDALRRVAGYPEE  
ST

NYAQYMSRAYAEFLGEDPGSGTDARPSLFWRLAGLLASSGFAFVNAHAHAIRLSDLLGFLAHSRVL  
AG

LAARGAAGCAADSVFLNVSVLDPAARLRLEARLGHVAAILEREQSLAAHALGYQLAFVLDSPPAAYGA  
VA

20 PSAARLIDALYAEFLGGRALTAPMVRRALFYATAVLRAPFLAGAPSAEQRRERARRGLLITTALCTSDV  
AA

ATHADLRAALARTDHQKNLFWLPDHFSPCAASLRFDLAEGGFILDALAMATRSDIPADVMAQQTRGVA  
SV

25 LTRWAHYNALIRAFVPEATHQCSGSPSHNAEPRILVPITHNASYVVVTHTPLRGIGYKLTGVDVRRPLF  
IT

YLTATCEGHAREIEPKRLVRTENRRDLGLVGAVFLRYTPAGEVMSVLLVDTDATQQQLAQGPVAGTPN  
VF

SSDVPSVALLLFPNGTVIHLLAFDTLPATIAPGFLAASALGVVMITAALAGILRVVRTCVPFWRRE  
\*

30

Gene matched: gi|138315|sp|P06477|VGLH\_HSV11

Gene name: GLYCOPROTEIN H PRECURSOR

35 [SEQ ID NO:237]

ORF # = 24 from Contig 15

ORF start site = 39188

ORF end site = 38058

ORF sequence:

40 MASHAGQQHAPAFGQAARASGPTDGRAASRPSHRQASEARGDPELPTLLRVYIDGPHGVGKTTTSAQ  
LM

EALGPRDNIVYVPEPMTYWQVLGASETLTNIYNTQHRLDRGEISAGEAAVVM TSAQITMSTPYAATDA  
VL

APHIGGEAVGPQAPPPALTLVFDHRPIASLLCYPARYLMGSMT PQAVLAFVALMPPTAPGTNLVLGV

45 LP

EAEHADRLARRQRPGERLDLAMLSAIRRVYDLLANTVRYLQRGGRWREDWGRLTGVAATPRPDPEAG  
AG  
SLPRIEDTLFALFRVPELLAPNGDLYHIFAWVLDVLADRLPMHLFVLDYDQSPVGCRDALLRLTAGM  
IP

5 TRVTTAGSIAEIRDLARTFAREVGGV\*

Gene matched: gi|125438|sp|P04407|KITH\_HSV23

Gene name: THYMIDINE KINASE

10

[SEQ ID NO:238]

ORF # = 25 from Contig 15

ORF start site = 39090

15 ORF end site = 39935

ORF sequence:

MARTGRRAAVGRPARTSSLTERRRVLLAGVRSHTRFYKAFAREVREFNATRICGTLTLMSSLSQGRS  
LF

20 EATRVTLICEVDLGP RPDCICVFEFANDKTLGGVCVILELKTCKSISSGDTASKREQRTTGMKQLRH  
SL

KLQSLAPPGDKVVYLCPILVFVAQRTLRSRVTRLVPQKISGNITAAVRMLQSLSTYAVPPEPQTRR  
SR

RRVAATARPQRPPSPTRDPEGTAGHPAPPESDPSPGVGVAAEGGGVLQKIAALFCVPVAAKSRPRT  
KT

25 E\*

Gene matched: gi|136854|sp|P10208|UL24\_HSV11

Gene name: PROTEIN UL24

30

[SEQ ID NO:239]

ORF # = 26 from Contig 15

ORF start site = 40216

ORF end site = 41973

35 ORF sequence:

MDPYYPFDALDVWEHRRFIVADSRSFITPEFPRDFWMLPVFNIPRETAAERA AVLQAQRTAAAAALEN  
AA

LQAAELPVDIERRIRPIEQQVHHIADALEALETAAAAEEADAARDAEARGEAGDGAAPSPTAGPAA  
AE

40 MEVQIVRNDPPLRYDTNLPVDLLHVMYAGRGAGSSGVVFGTWYRTIQERTIADFPLTTRSADFRDGR  
MS

KTFMTALVLSLQSCGRLYVGQRHYSAFECAVLCLYLLYRTTHESSPDRDRAPVAFGDLLARLPYRLAR  
LA

AVIGDESGRPQYRYRDDKLPKQFAAAGGRYEHGALATHVVIATLVRHGVLPAAPGDVPRDTSTRVNP

45 DD

VAHRDDVNRAAAFLARGHNFLWEDQTLRATANTITALAVLRRLLANGNVYADRLDNRLQLGMLIP  
GA  
VPAAEAIARGASGLDSGAIKSGDNNLEALCVNYVLPYQADPTVELTQLFPGLAALCLDAQAGRPLAST  
RR  
5 VVDMSSGARQAALVRLTALELINRTRTNTTPVGEIINAHDAIGIQYEQGLGLLAQQARIGLASNAKRF  
AT  
FNVGSDYDLLYFLCLGFIPQYLSVA\*

10 Gene matched: gi|136863|sp|P10209|UL25\_HSV11  
Gene name: VIRION PROTEIN UL25

[SEQ ID NO:240]  
ORF # = 27 from Contig 15  
15 ORF start site = 42206  
ORF end site = 44179  
ORF sequence:  
MASAEMRERLEAPLPDRAVPPIYVAGFLALYDSGDPGELALDPDTVRAALPPENPLPINVDHRARCEVG  
RV  
20 LAVVNDPRGPFVGLIACVQLERVLETAASAAIFERRGPALSREERLLYLITNYLPSVSLSTKRRGDE  
VP  
PDRTLFAHVALCAIGRRLGTIVTYDTSLDAAIAPFRHLDPATREGVRREAAEAELALAGRTWAPGVEA  
LT  
HTLLSTAVNNMLRDRWSLVAERRRQAGIAGHTYLQASEKFKIWAESAPAPERGYKTGAPGAMDTSP  
25 AA  
SVPAPQVAVRARQVASSSSSSSFAPADMNPVSASGAPAPPPPGDGSYLWIPAFHYNQLVTGQSAPH  
HP  
PLTACGLPAAGTVAYGHPGAGPSPHYPPPAHPYPGMLFAGPSPLEAQIAALVGAIADRQAGGLPAA  
AG  
30 DHGIRGSAKRRRHEVEQPEYDCGRDEPDRDFPYYPGEARPEPRPVDSRRAARQASGPHEITITALVGAV  
TS  
LQQELAHMRARTHAPYGPYPVGPYHHPHADTETPAQPPRYPAEAVYLPFPHIAPPGPPLSGAVPPPS  
YP  
PVAVTGPGAPPLHQPSPAHAHPPPPPGPTPPPAASLPQPEAPGAEAGALVNASSAAHVKRGHGPGRR  
35 SV  
CVTDDGVPLTRLQDPDLGGVCVFIYFK\*

Gene matched: gi|139233|sp|P10210|VP40\_HSV11  
Gene name: CAPSID PROTEIN P40 (VIRION S

40

[SEQ ID NO:241]  
ORF # = 28 from Contig 15  
ORF start site = 47298  
45 ORF end site = 44584

ORF sequence:

MRGGGLICALVVGALVAAVASAAPAAPASGGVAATVAANGGPASRPPVPSPATTKARKRKT  
 PP  
 KRPEATPPPDANATVAAGHATLRAHLREIKVENADAQFYVCPPTGATVVQFEQPRRCPTRPEGQNYT  
 5 EG  
 IAVVFKENIAPYKFKATMYKDVTVSQVWFGHRYSQFMGIFEDRAPVPFEEVIDKINAKGVCRSTAKY  
 VR  
 NNMETTAFHRDDHETDMELKPAKVATRTSRGWHTTDLKYNPSRVEAFHRYGTTVNCIVEEVDARSVYP  
 YD  
 10 EFVLATGDFVYMSPFYGYREGSHTSHTSYAADRFKQVDGFYARDLTTKARATSPTTRNLLTTPKFTVA  
 WD  
 WVPKRPAVCTMTKWQEVDEMLRAEYGGSSFRFSSDAISTTFTTNLTQYSLSRVDLGDCIGRDAREIDR  
 MF  
 ARKYNATHIKVGQPQYYLATGGFLIAYQPLLSNTLAELYVREYMREQDRKPRNATPAPLREAPSANAS  
 15 VE  
 RIKTTSSIEFARLQFTYNIQRHVNDMLGRIAVAWCELQNHETLWNEARKLNPNAIASATVGRRVSA  
 RM  
 LGDVMVAVSTCVVAPDNVIVQNSMRVSSRPGTCYSRPLVSFRYEDQGPLIEGQLGENNELRLTRDALE  
 PC  
 20 TVGHRRYFIFGGGYVYFEEYAYSHQLSRADVTTVSTFIDLNITMLEDHEFVPLEVYTRHEIKDSGLLD  
 YT  
 EVQRRNQLHDLRFADIDTVIRADANAAMFAGLCAFFEGMGDLGRAVGKVVMGVVGGVVSAVSGVSSFM  
 SN  
 PFGALAVGLLVLAGLVAAFFAFRYVLQLQRNPMKALYPLTTKELKTSDPGGVGEGEGEAGEGGGFDEA  
 25 KL  
 AEAREMIRYMALVSAMERTEHKARKKCTSALLSSKVTNMVLRKRNKARYSPLHNEDEAGDEDEL\*

Gene matched: gi|138198|sp|P06763|VGLB\_HSV23

30 Gene name: GLYCOPROTEIN B PRECURSOR.gi|

[SEQ ID NO:242]

ORF # = 29 from Contig 15

35 ORF start site = 47122

ORF end site = 47338

ORF sequence:

VVAGLGTGGGREAGPPFAATVAATPPEARGAAGAAGAATAATSAPTTSQIKPPPRMAGLRGRVAP  
 AA  
 40 R\*

Gene matched: gi|729379|sp|P39055|DYN1\_CAEEL

Gene name: DYNAMIN.gi|456286 (L29031) d

45



[SEQ ID NO:243]  
 ORF # = 30 from Contig 15  
 ORF start site = 49662  
 5 ORF end site = 47305  
 ORF sequence:  
 MAAAPPAAVSEPTAARQKLLALLGQVQTYVFQLELLRRCDPQIGLGKLAQLKLNALQVRVLRRLRPG  
 LE  
 AQAAFLTPLSVTLELLLEYAWREGERLLGHLETFATTGDVSAFFTETMGLARPCPYHQIRLETYGG  
 10 DV  
 RMELCFLDHVENFLKQLNYCHLITPPSGATAALERVREFMVAAVGSGLIVPELSDPSHPCAVCFEEL  
 CV  
 TANQGATIAARRLADRICNHVTQQAQVRLDANELRRYLPHAAGLSDAARARALCVLDQALARTAAGGGA  
 RA  
 15 GPPPADSSSVREEADALLEAHDVFQATTPGLYAISELRFWLASGDRARHSTMDAFADNLNALAQRELQ  
 QE  
 TAAVAVELALFGRRAEHFDRAFGGHLAALDMVDALIIGGQATSPDDQIEALIRACYDHHLTTPLLRRL  
 VS  
 PEQCDEEALRRVLARLGAGGATGGAEETEPRAAAEEGRRRGAGTPASEDGERGPEPGAQGPESWGDI  
 20 AT  
 RAAADVPERRRLYADRLTKRSLASLGRCVREQRGELEKMLRVSVHGEVLPATFAAVANGFAARARFCA  
 LT  
 AGAGTVIDNRAAPGVFDAHRFMRASLLRHQVDPALLPSITHRFFELVNGPLFDHSTHSFAQPPNTALY  
 YS  
 25 VENUGLLPHLKEELARFIMGAGGSGADWAVSEFQKFYCFDGVSGITPTQRAAWRYIRELIIATTLFAS  
 VY  
 RCGELELRRPDCSRPTSEGLYRYPGVYLTYNDCPLVAIVESGPDGCIGPRSVVYDRDVFSLYSV  
 LQ  
 HLAAPRLAGGGSDAPP\*

30

Gene matched: gi|124088|sp|P10212|P RTP\_HSV11

Gene name: PROCESSING AND TRANSPORT PRO

35

[SEQ ID NO:244]  
 ORF # = 31 from Contig 15  
 ORF start site = 51666  
 ORF end site = 50035  
 40 ORF sequence:  
 MSLSLDPYTCGPCPLLQLLARRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFPVLRRLRVMDLFNN  
 GF  
 LSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFPKELRVKSRVLFAGASANASEAAK  
 AR

VASLQSA YQKPKRVDILLGPLGFLKQFHAVIFPNGKPPGSNQPNPQWFTALQRNQLPARLLSRED  
IE  
TIAFIKRFSLDYGAINFINLAPNNVSELAMYYMANQILRYCDHSTYFINTLTAVIAGSRRPPGVQAAA  
AW  
5 APQGGAGLEAGARALMDSLD AHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAG  
NW  
ASLLGGKNACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHEHSLCEQLRGIIAEGGA AVASSVFVA  
TV  
KSLGPRTQQLQIEDWLALLEDEYLSEEMMEFTTRALERGHGEWSTDAALEVAHEAEALVSQ LGAAGEV  
10 FN  
FGDFGDEDDHAASFGLAAAAGAAGVARKRAFHGDDPFGEGPPEKKDLTLDML\*

Gene matched: gi|544182|sp|P36384|DNBI\_HSV2  
15 Gene name: MAJOR DNA-BINDING PROTEIN (IN

[SEQ ID NO:245]  
ORF # = 32 from Contig 15  
20 ORF start site = 53575  
ORF end site = 51701  
ORF sequence:  
MDTKPKTTTTTKVPPGPMGYVYGRACPAEGLELLSLSARS GDADVAVAPLIVGLTVESGF EANVA AV  
VG  
25 SRTTGLGGTAVSLKLMPSHYSPSVYVFHGGRHLPSTQAPNLTRLCERARRHFGFSDYAPRPCDLKHE  
TT  
GDALCERLGLDPDRALLYLVITEGFREAVCISNTFLHLGGMDKVTIGDAEVHRIPVYPLQMFMPDFSR  
VI  
ADPFNCNHRSIGENFNYPFPFNRPLARLLFEAVVGPA AVALRARNVDAVARAAAH LAFDENHEGAAL  
30 PA  
DITFTAFEASQGKPQRGARDAGNKGPGAGFEQRLASVMAGDAALALE SIVSMAVFDEPPPDITTWPLL  
EG  
QETPAARAGAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSKPSFYRFFLVPGTHVAAN  
PQ  
35 LDREGHVVPGYEGRPTAPLVGGTQEFAGEH LAMLCGFSPALLAKMLFYLERCDGGVIVGRQEMDVFRY  
VA  
DSGQTDVPCNLCTFETR HACAHTTLMRLRARHPKFASAARGAIGVFGTMNSAYS DCDVLGNYAAFSAL  
KR  
ADGSENTRTIMQETYRAATERVMAELEALQYVDQAVPTALGRLETTIIGTREALHTV VNNIKQLV\*  
40

Gene matched: >gi|544182|sp|P36384|DNBI\_HSV2  
Gene name: MAJOR DNA-BINDING PROTEIN (IN

45

[SEQ ID NO:246]  
ORF # = 33 from Contig 15  
ORF start site = 54393  
ORF end site = 58115  
5 ORF sequence:  
MFCAAGGPTSPGGKSAARAASGFFAPHNPRGATQTAPPPCRRQNFYNPHLAQTGTQPKAPGPAQRHTY  
YS  
ECDEFRFIAPRSLDEDAPAEQRTGVHDGRLRRAPKVYCGGDERDVLRVGPEGFWPRRLRLWGGADHAP  
EG  
10 FDPTVTVFHVVDILEHVEHAYSMRAAQLHERFMDAITPAGTVITLLGLTPEGHRVAVHVYGTROYFYM  
NK  
AEVDRHLQCRAPRDLCLERLAAALRESPGASFRGISADHFEEVVERADVYYYETRPTLYYRVFVRSR  
AL  
AYLCDNFCAIRKYEGGVDAATTRFILDNPGFVTFGWYRLKPGRCNAPAQPRPPTAFGTSSDVEFNCTA  
15 DN  
LAVEGAMCDLPAYKLMCFDIECKAGGEDELAFPAERPDLVIQISCLLYDLSTTALEHILLFSLGSC  
DL  
PESHLSDLASRGLPAPVVLEFDSEFEMLLAFMTFVKQYGPEFVTGYNIINFDPFVLTKLTEIYKVPL  
DG  
20 YGRMNGRGVFRVWDIGQSHFQKRSKIKVNGMVNIDMYGIITDKVKLSSYKLNVAEAVLKDKKKDLSY  
RD  
IPAYYASGPAQRGVIGEYCVQDSLLVGQLFFKFLPHLELSAVARLAGINITRTIYDGGQIRVFTCLLR  
LA  
GQKGFILPDTQGRFRGLDKEAPKRPAVPRGEGERP GDNGDEDKDDDEDGDEDEREEVARETGGRH  
25 VG  
YQGARVLDPTSGFHVDPVVVDFASLYPSIIQAHNLCFSTLSLRPEAVAHLEADR DYLEIEVGGRRLF  
FV  
KAHVRESLLSILLRDWLAMRKQIRSRI PQSTPEEAVLLDKQQAIAKVVCNSVYGFTGVQHGLLPCLHV  
AA  
30 TVTTIGREMLLATRAYVHARWAEFDQLLADFPEAAGMRAPGPYSMRIIYGDTSIFVLCRGLTAAGLV  
AM  
GDKMASHISRALFLPPIKLECEKTFTKLLLI AKKKYIGVICGGKMLIKGVDLVRKNNCAFINRTSRAL  
VD  
LLFYDDTVSGAAAALAEPAEEWLRPLPEGLQAFGAVLVDHRRITDPERDIQDFVLTAELSRHPRA  
35 YT  
NKRLAHLTVYVKLMARRAQVPSIKDRIPYVIVAQTREVEETVARLAALRELDAAAPGDEPAPPAALPS  
PA  
KRPRETPSHADPPGGASKPRKLLVSELAEDPGYAIARGVPLNTDYF SHLLGAACVTFKALFGNNAKI  
TE  
40 SLLKRFIPETWHPPDDVAARLRAAGFGPAGAGATAEETRMLHRAFDTLA\*

Gene matched: gi|118882|sp|P07918|DPOL\_HSV21

Gene name: DNA POLYMERASE

45

[SEQ ID NO:247]  
ORF # = 34 from Contig 15  
ORF start site = 58977  
5 ORF end site = 58060  
ORF sequence:  
MYDIAPRRSGSRPGPGRDKTRRRSRFSAAGNPGVERRASRKSLSHARRLELCLHERRRYRGFFAALA  
QT  
PSEEIAIVRSLSVPLVKTTFPVSLPFSLDQTVADNCLTLSGMGYYLGIGGCCPACSAGDGRLATVSREA  
10 LI  
LAFVQQINTIFEHRTFLASLVVLADRHSTPLQDLLADTLGQPELFFVHTILRGGGACDPRFLFYPDPT  
YG  
GHMLYVIFPGTSAHLHYRLIDRMLTACPGYRFAAHVWQSTFVLVVRNAEKPADAEIPTVSAADIYCK  
MR  
15 DISFDGGLMLEYQRLYATFDEFPPP\*

Gene matched: gi|136875|sp|P10215|UL31\_HSV11  
Gene name: PROTEIN UL31

20 [SEQ ID NO:248]  
ORF # = 35 from Contig 15  
ORF start site = 60760  
ORF end site = 58970  
25 ORF sequence:  
MATSAPGVPSSAAVREESPGSSWKEGAFFERPYVAFDPDLLALNEALCAELLAACHVVGVPASPALDED  
VE  
SDVAPAPPRPRGAAREASGGRGPGSARGPPADPTAEGLLDTGPFAAASVDTFALDRPCLVCRTIELYK  
QA  
30 YRLSPQWVADYAFCLCAKCLGAPHCAASIFVAAFEFVYVMDHHFLRTKKATLVGSFARFALTINDIHRH  
FF  
LHCCFRTDGGVPGRHAQKQPRPTSPGAAKVQYSNYSFLAQSATRALIGTLASGGDDGAGAGGSGTQ  
PS  
LTTALMNWKDCARLLDCTEGKRGGDSCCTRAAARNGEFEAAAGALAQQGEPETWAYADLILLLLAGT  
35 PA  
VWESGPRLRAAADARRAAVSESWEAHRGARMRDAAPRFAQFAEPKAQPDLDLGPLMATVLKHGRGRGR  
TG  
GECLLCNLLLVRAYWLMRRLRASVVRYSENNTSLFDCIVPVVDQLEADPEAQPGDGGRFVSLLRAAG  
PE  
40 AIFKHMFCDPMCAITEMEVDPWVLFHPRADHRDELQLHKAKLACGNEFEGRVCIALRALIYTFKTYQ  
VF  
VPKPTALATFVREAGALLRRHSISLLSLEHTLCTYV\*

Gene matched: gi|136879|sp|P10216|UL32\_HSV11  
45 Gene name: PROBABLE MAJOR ENVELOPE GLYC

[SEQ ID NO:249]  
ORF # = 36 from Contig 15  
5 ORF start site = 60759  
ORF end site = 61151  
ORF sequence:  
MAGRAGRTRPRTLDAIPDCALRSQTLES LDARYVSRDGAGDAAVWFEDMTPAELEVIFPTTDAKLNY  
LS  
10 RTQRLASLLTYAGPIKAPDGPAPHTQDTACVHGELLARKRERFAAVINRFLDLHQILRG\*

Gene matched: gi|136883|sp|P10217|UL33\_HSV11  
Gene name: UL33  
15

[SEQ ID NO:250]  
ORF # = 37 from Contig 15  
ORF start site = 61241  
20 ORF end site = 62071  
ORF sequence:  
MAGMGKPYGGRPGDAFEGLVQIRILIVPTTLRGGGGESGPYSPSNPPSRCAFQFHGQDGSDEAFPIEY  
VL  
RLMNDWADVPCNPYLRVQNTGVS VLFQGGFFNRPHGAPGGAITAEQTNVILHSTETTGLSLGDLDDVKG  
25 RL  
GLDARPMASMWISCFVRMPRVQLAFRFMGPEDAVRTRILCRAAEQALARRRRSRRSQDDYGAVAVA  
AA  
HHSSGAPGPGVAASGPPAPPGRGPAPWHQAVQLFRAPRPGPPALLLVAGLFLGAAIWWAVGARL\*

30

Gene matched: gi|136888|sp|P10218|UL34\_HSV11  
Gene name: VIRION PROTEIN UL34

35 [SEQ ID NO:251]  
ORF # = 38 from Contig 15  
ORF start site = 62183  
ORF end site = 62521  
ORF sequence:  
40 MAAPQFHRPSTITADNVRALGMRGLVLATNNAQFIMDNSYPHPHGTQGA VREFLRGQAAALTDLGVTH  
AN  
NTFAPQPMFAGDAAAEWLRPSFGLKRTYSPFVVVRDPKTPSTP\*

45 Gene matched: gi|139196|sp|P10219|VP26\_HSV11

Gene name: CAPSID PROTEIN VP26

ID[SEQ ID NO:252]

5 ORF # = 39 from Contig 15  
ORF start site = 72047  
ORF end site = 62688  
ORF sequence:  
MIPAAALPHPTMKRQGDRDIVVTGVRNQFATDLEPGGSVSCMRSSLSFLSLLFDVGPRDVLSAEAEIGC  
10 LV  
EGGEWTRAAAGSGPPRMCSIIELPNFLEYPAARGGLRCVFSRVYGEVGFGEPTAGLLETQCPAHTFF  
AG  
PWAMRPLSYTLTIGPLGMGLYRDGDTAYLFDPHGLPAGTPAFIAKVRAGDVYPYLTYAHDRPKVRW  
AG  
15 AMVFFVPSGPGAVAPADLTAAALHLYGASETYLQDEPFVERRVAITHPLRGEIGGLGALFVGVPVPRGD  
GE  
GSGPVVPALPAPTHVQTPRADRPPEAPRGASGPPNTPQAGHPNRPPDDVWAAALEGTPPAKPSAPDAA  
AS  
GPPHAAPFPQTPAGDAAEEAEDLRVLEVGA VPGVGRHRYSTGLPKRRRPTWTTPSSVEDLTSGERPA  
20 PK  
APPAKAKKKSAPKKKAPVAAEVPASSPTPIAATVPPAPDTPPQSGQGGGDDGPASPSSPSVLETGAR  
RP  
PEPPGADLAQLFEVHPNVAATAVRLAARDAALAREVAACSQLTINALRSPYPAHPGLELCVIFFFER  
VL  
25 AFLIENGARTHTQAGVAGPAAALLDFTLRMLPRKTAVGDFLASTRMSLADVAHRPLIQHVLDENSQI  
GR  
LALAKLVLVARDVIRETDAFYGDLADLDLQLRAAPPANLYARLGEWLLERSRAHPNTLFAPATPTHPE  
PL  
LHRIQALAQFARGEEMRVEAEAREMREALDALARGVDSVSQRAGPLTVMPVPAAPGAGGRAPCPPALG  
30 PE  
AIQARLEDVRIQARRAIESAIKEYFHRGAVYSAKALQASDSHDCRFHVASAAVPMVQLLES LPAFDQ  
HT  
RDVAQRAALPPPPPLATSPQAILLRDLLQRGQTL DAPEDLAAWLSVLTDAATQGLIERKPLEELARSI  
HG  
35 INDQQARRSSGLAELQRFDALDAALAQLDSDAAFPATGPAPYVDGGGLSPEATRMAEDALRQARAM  
EA  
AKMTAELAPEARSRLRERAHAEAMLNDARERAKVAHDAREKFLHKLQGVLRPLPDFVGLKACPAVLA  
TL  
RASLPAGWTDLADAVRGPPPEVTAALRADLWGLLGQYREALEHPTPD TATALAGLHPAFVVVLKTLFA  
40 DA  
PETPVLVQFFSDHPTIAKAVSNAINAGSAAVATASPAATVDAVRAHGALADAVSALGAAARDPAS  
LS  
FLAALADSAAGYVKATRLALEARGAIDELTTLGSAADLVVQARRACAQPEGDHAALIDAAARATTAA  
RE

SLAGHEAGFGGLLHAEGTAGDHSPSGRALQELGKVGATRRRADELEAAVADLTAKMAAQRARGSSER  
WA  
AGVEAALDRVENRAEFDVVELRRLQALAGTHGYNPRDFRKRAEQALANAEEAVTLALDTAFANPYTP  
EN  
5 QRHPMLPPLAAIHRLGWSAAFHAAAETYADMFRVDAEPLARLLRIAEGLLEMAQAGDGFIDYHEAVGR  
LA  
DDMTSVPGLRRYVFFQHGADYVELRDRLDAIRADVHRALGGVPLDLAAAEEQISAARNDEATAEL  
VR  
TGVTLPSPSEDALVACAAALERVQSPVKNTAYAEYVAFVTRQDTAETKDAVVRKQQRAEATERVMA  
10 GL  
REALAARERRAQIEAEGLANLKTMLKVAVPATVAKTLDQARSVAEADQVEVLLDQTEKTRELDVPA  
VI  
WLEHAQRTFETHPLSAARGDGPGLARHAGRLGALFDTRRRVDALRRSLEEAEAEWDEVWGRFGRVRG  
GA  
15 WKSPEGFRAMHEQLRALQDTTNTVSGLRAQPAYERLSARYQGVLGAKGAERAEAVEELGARVTKHTAL  
CA  
RLRDEVVRRVPWEMNFDALGRLLAEFDAAAADLAPWAVEEFRGARELIQRRMGLYSAYARAGGQTGAG  
AA  
AAPAPLLVDLRALDARARASSPEGHEVDPQLRRRGEAYLRAGGDPGPLVLEAVSALDLPFATSFL  
20 AP  
DGTPLQYALCFPAVTDKLGALLMRPEAACVRPPLPTDVLESAPTVTAMYVLTVVNRLQLALSDAQAN  
FQ  
LFGFRVRRHQATWGASMDAAELYVALVATTLTREFGCRWAQLGWASGAAAPRPPGPRGSQRHCVA  
NE  
25 NDVLVALVAGVPEHIYNFWRLDLVRQHEYMHLTLERAFEDAAESMLFVQRLTPHPDARIRVLPFLDG  
GP  
PTRGLLFGTRLADWRRGKLGSETDPLAPWRSALGLGTQRRDAPALGKLSPAQALAAVSVLGRMCLPSAA  
LA  
ALWTCMFDDYTEYDSFDALLAARLESQTLGPAGGREASLPEAPHALYRPTGQHVAVLAAATHRTPA  
30 AR  
VTAMDVLVAVLLGAPVVVALRNTTAFSRESELELCLTLFDSRPGPDALRDVVSSDIETWAVGLLH  
TD  
LNPIENACLAALPRLSALIAERPLADGPPCLVLVDISMTPVAVLWEAPEPPGPPDVRVFGSEATEEL  
PF  
35 VATAGDVLAAASAADADPFFARAILGRPFDA\$LLTGELFPGHPVYQRPLADEAGPSAPTAARDPRDLA  
GD  
GGSGPEDPAAPPARQADPGVLAPTLLTDATTGEPVPPRMWAWIHGLEELASDDAGGPTPNPAPALLPP  
PA  
TDQSVPTSQYAPRPIGPAATARETRPSVPPQONTGRVPVAPRDDPRSPPTPSPPADAALPPPAFSGS  
40 AA  
AFSAAVPRVRRSRTRAKSRAPRASAPPEGWRPPALPAPVAPVAASARPPDQPTPESAPPAWVSALP  
LP  
PGPASARGAFPAPTLAPIPPPPAEGAVAPGDDRRRGRRQTTAGPSPTPPRGPAAGPPRRLTRPAVASL  
SA

SLNSLPSPRDPADHAAAVSAAAAVPPSPGLAPPTSAVQTSPPPLAPGPFVAPSEPLCGWVVPGGPVAR  
 RP  
 PPQSPATKPAARTRIRARSVPQPPLPQPPLPQPPLPQPPLPQPPLPQPPLPQPPLPQPPLPQPPLPQP  
 PL  
 5 PQPPLPQPPLPQSRDSVPTPESPTHNTNTHLPVSAVTSWASSLALHVDSAPPPASLLQTLHISSDDEHS  
 DA  
 DSLRFSDSDDEALDPLPPEPHLPPEDEPPGFLAADHLQSPHSQFGPLPVQANAVLSRRYVVRSTGRSA  
 LA  
 VLIRACRRIQQQLQRTRRALFQRSNAVLTSLHHVRMLLG\*  
 10  
 Gene matched: gi|135576|sp|P10220|TEGU\_HSV11  
 Gene name: LARGE TEGUMENT PROTEIN (VIRI  
 15 [SEQ ID NO:253]  
 ORF # = 40 from Contig 15  
 ORF start site = 75699  
 ORF end site = 72355  
 ORF sequence:  
 20 MSDSALQVPAPAGMTPPSAPPPNGPLQVLLGSLTNLRRPPSPSSEPADEPAFLSAAKLRAATAAF  
 LL  
 SGAAVGPAEARACWHPLLEQLCALHRAHGLPETALLAENLPGLLVHRMAVALPETPEAAFREMDVIKD  
 TV  
 LAITGSDTTHALEAAGLRTTAALGPVRVRQCAVEWIDRWRTVTQSLAMNPRTSLEALGEMSLKMSPV  
 25 PL  
 GQPGANLTPAYSLLFPSPIVQEGLRFLALVSNWVTLFSAHLQRIDDAALTPLTRALFTLALVDEYLT  
 TP  
 DRGAVVPPLLAQFQHTVREIDPAIMIPPLEATKMVRSREEVRVSTALSRVSPRSACAPPGTLMARVR  
 TD  
 30 AAVFDPDVFPFLSASALAFRPAVTGLLQLGEPPSAGAQQRLALLQQTWALVQNSNSPSVVINTLTDA  
 GF  
 TPAHCTQYISALEGFLVAGVPARTPPGHGLSEIQQQLFGCIALAGANVFLAREYGHYAGYVKTFRRIQ  
 GA  
 SEHTHGRICEAVGLSGGVLSQTLARIMGPAVPTEHLASLRRTLVEGEFETAERRFSAGQPSLLRETALI  
 35 WL  
 DVGQTHWDLTPTTPATPLSALLPVGPPSHAPSVHLAAATKIRFPALGHPNVLADPGFVPYVLALV  
 VG  
 DALRATCNAAAYLPRPIEFALRVLAWARDFGLGYLPTVEGHRTKLALITLLEPATRAGVGPTMQMADN  
 IE  
 40 QLLRELYVIARGAVEQLRPAVQLPPPQPPEVGSSLLISMYALAARGVLQELAERADPLVRQLEDAIV  
 LL  
 RLHMRTLAAFFECRFESDGHRLYAVVADAHERLGPWRPEAMGDAVSQYCGMYHDAKRALVASLAGLRS  
 VV  
 TETTAHLGVCDLAAQVSHEGNVLAVVRREIHGFLAIVSGIHARASKLMSGDQVPGFCYMSQFLARWR  
 45 RL



SAGYQAARAATGPVERAEFVQELHDTWKGLQTERALVVAPFASSADQRTAAIQEVMAHATEDAPPSPA  
AD  
LVVLNRHDLGAWGDYSLGPLGQPTVVPDSVDLSPQGLAATLSMDWLLINELLQVTDGVFRASAFRPS  
AG  
5 PEAPGDLEAQDAGGSTPEPTTPGPQDTQARAPSTRPAGRETVWPWNTPVEDDEMPQETPPVHP\*

Gene matched: gi|136894|sp|P10221|V120\_HSV11

Gene name: CAPSID ASSEMBLY PROTEIN UL37

10

[SEQ ID NO:254]

ORF # = 42 from Contig 15

ORF start site = 78158

15 ORF end site = 81592

ORF sequence:

MANRPAASALAGARSPSERQEPREPEVAPPGDHFVCRKVSVMVLSSDPPGPAAYRISDSSFVQCGS  
NC  
SMIIDGDVARGHLRDLEGATSTGAFVAISNVAAGDGRTAVVALGGTSGPSATTSVGTQTSGEFLHGN  
20 PR  
TPEPQGPQAVPPPPPPFPWGHECCARRDARGGAEKDVGAAESWSDGPSSDSETEDSDSSDEDTGSGS  
ET  
LSRSSSIWAAGATDDDDSDSDSRSDSVQPDVVRRRWSDGPAFVAFKPRRPGDSPGNPGLGAGTG  
GS  
25 ATDPRASADSDSAAHAAAPQADVAPVLDSQPTVGTDPGYVPVPLELTPENAEAVARFLGDAVDREPALM  
LE  
YFCRCAREESKRVPRTFGSAPRLTEDDFGLLNYALAEMRRLCLDLPPVPPNAYTPYHLREYATRLVN  
GF  
KPLVRRSARLYRILGILVHLRIRTREASFEEWMRSKEVDLDFGLTERLREHEAQLMILAQALNPYDCL  
30 IH  
STPNTLVERGLQSALKYEEFYLRFGGHYMESVFQMYTRIAGFLACRATRGMRHIALGRQGSWEMFK  
FF  
FHRLYDHQIVPSTPAMNLGTRNYYTSSCYLVNPQATTNQTALRAITGNVSAILARNGGIGLCMQAFN  
DA  
35 SPGTASIMPALKVLDLVAAHNKQSTRPTGACVYLEPWHSDVRAVLRMKGVLAGEEAQRCDNIFSALW  
MP  
DLFFKRLIRHLDGEKNVTWSLFDRTSMSLADFHGEEFEKLYEHLEAMGFGETIPIQDLAYAIVRSA  
TT  
GSPFIMFKDAVNRHYIYDTQGAAIAGSNLCTEIVHPSSKRSSGVCNLGSVNLARCVSRRTDFGMLRD  
40 AV  
QACVLMVNIMIDSTLQPTPQCARGHDNLRSMGIGMQGLHTACLKMGDLLESAEFRDLNTHIAEVMLLA  
AM  
KTSNALCVRGARPFSHFKRSMYRAGRHFHWERFSNASPRYEGEWMLRQSMMKHGLRNSQFIALMPTAA  
SA

QISDVSEGFAPLFTNLFSKVTRDGETLRPNTLLKELERTFGGKRLLDAMDGLEAKQWSVAQALPCLD  
PA  
HPLRRFKTAFDYDQELLIDLCADRAPYVDHSQSMTLYVTEKADGTLASTLVRLLVHAYKRGLKTGMY  
YC

5 KVRKATNSGVFAGDDNIVCTSCAL\*

Gene matched: gi|1710385|sp|P09853|RIR1\_HSV23  
Gene name: RIBONUCLEOSIDE-DIPHOSPHATE

10

[SEQ ID NO:255]  
ORF # = 43 from Contig 15  
ORF start site = 81665  
ORF end site = 82658

15 ORF sequence:

MDPAVSPASTDPLDTHASGAGAAPIVCPTPERYFYTSQCPDINHLRSLILNRWLETFLVFGDEED  
VS  
KLSEGELGFYRFLFAFLSAADDLV TENLGGLSGLFEQKDILHYVVEQECIEVVHSRVYNIQVLVFN  
ND

20 QARRAYVARTINHPAIRVKVDWLEARVRECDSEKFIEMILIEGVFFAASF AAIAYLR TNLLRVTC  
QS  
NDLISRDEAVHTTASCIYNNYLGGHAKPEARVYRLFREAVDIEIGFIRSQAPTDSSILSPGALAAI  
EN  
YVRF SADRLGLIHMQLYSAPAPDASFPLSLMSTDKHTNFFECRSTSYAGAVVNDL\*

25

Gene matched: gi|132624|sp|P03174|RIR2\_HSV23  
Gene name: RIBONUCLEOSIDE-DIPHOSPHATE R

30

[SEQ ID NO:256]  
ORF # = 44 from Contig 15  
ORF start site = 84014  
ORF end site = 82941

35 ORF sequence:

MRRRGHAFAPGDRGTRAAGPGPAAPWGAPSKPALRLAHLFCIRVLRALGYAYINSGQLEADDACANLY  
HT  
NTVAYVHTTDTDLLMGCDIVLDISTGYIPTIHCRDLLQYFKMSYPQFLALFVRCHTDLHPNNTYASV  
ED

40 VLRECHWTAPSRSQARRAARRERANSRSLESMP TLTAAPVGLETRISWTEILAQQIAGEDDYEDPPL  
QP  
PDVAGGPRDGARSSSSEILTPELVQVPNAQRVAEHRGYVAGRRRHVIHDAPEALDWLPDPMTIAELV  
EH

45 RYVKYVISLISPKERGFWTLLKRLPIYQDLRDEDLARSIVTRHITAPDIADRFLAQLWAHAPPPAFYK  
DV

LAKFWDE\*

Gene matched: gi|549322|sp|P36699|VHS\_HSV2G

5 Gene name: VIRION HOST SHUTOFF PROTEIN

[SEQ ID NO:257]

ORF # = 45 from Contig 15

10 ORF start site = 84914

ORF end site = 86326

ORF sequence:

MAHLPGGAAAAPLSEDAIPSPRTEREDWPPCQIVLQGAELNGILQAFAPLRTSLDLSLLVVGDRGILV  
HN15 AIFGEQVFLPLDHSQFSRYRWGGPTAAFLSLVDQKRSLLSVFRANQYPDLRRVELTVTGQAPFRTL  
RIWTTASDGEAVELASETLMKRELTSFAVLLPQGDPDVQLRLTKPQLTKVVNAVGDDETAKPTTFELGPNG  
KF20 SVFNARTCVTFAAREEGASSSTSAQVQILTSALKKAGQAAANAKTVYGENTHRTFSVVVDDCSMRAVL  
RRLQVGGGTLKFFLTADVPSVCVTATGPNVSAVFLLPQVRVCLNLGRTPGSSTGSLASQDSRAGPTDS  
QDFSSEPDAGDRGAPEEEGLEQARVPPAFPEPPGTKRRHAGAEVVPADDAKRPKTGVPAAPTRAESP  
LS

25 ARYGPEAAEGGGDGGRYACYFRDLQTDASPSPLSAFRGPQRPYPYGFGLP\*

Gene matched: gi|136905|sp|P10226|VPAP\_HSV11

Gene name: DNA POLYMERASE PROCESSIVITY

30

[SEQ ID NO:258] = 15

ORF # = 48 from Contig 15

ORF start site = 89794

35 ORF end site = 90312

ORF sequence:

MAFRASGPAYQPLAPAASPARARVPAVAVIGVGAIVGAFALVAALVLVPPRSSWGLSPCDSGWQEFNA  
GC40 VAWDPTPVEHEQAVGGCSAPATLIPRAAAKHLAALTRVQAERSSGYWWVNGDGIRTCLRLVDSVSGID  
EF

CEELAIRICYPRSPGGFVRVFTSIRNALGLP\*

Gene matched: gi|136917|sp|P06483|UL45\_HSV23

Gene name: PROTEIN UL45 HOMOLOG (18 KD)

45

[SEQ ID NO:259]  
 ORF # = 49 from Contig 15  
 ORF start site = 92744  
 5 ORF end site = 90579  
 ORF sequence:  
 MQRARGASSRLARCLTPANLIRGANAGVPERRIFAGCLLPTPEGLLSAAVGVLQRADDLQPAFLT  
 GA  
 DRSVRLAARHNTVPESLIVDGLASDPHYDIRHYASAAKQALGEVELSGGQLSRAILAQYWKYLQTV  
 10 VP  
 SGLDIPDDPAGDCDPSLHVLLRPTLLPKLLVRAPFKSGAAAAYAAVAGLRDAAHRLQYMFMRPA  
 DP  
 SRPSTDALRLSEFLAYVSVLYHWASWMLWTADKYVCRRLGPADRRFVALSGSLEAPAETFARHLDRG  
 PS  
 15 GTTGSMQCMALRAAVSDVLGHLTRLAHLWETGKRSGGTYGIVDAIVSTVEVLSIVHHHAQYIINATLT  
 GY  
 VVWASDSLNNEYLRAAVDSQERFCRTAAPLFPTMTAPSWARMELSIKSWFGAALAPDLLRSGTSPSPHY  
 ES  
 ILRLAASGPPGGRGAVGGSCRDKIQRTRDNAPPPLPRARPHSTPAAPRRFRRHREDLPEPPHVDAAD  
 20 RG  
 PEPCAGRPATYYTHMAGAPPRLPPRNPAPPEQRPAARPLAAQREAAGVYDAVRTWGPDAEAEPDQM  
 EN  
 TYLLPDDDAAMPAGVGLGATPAADTTAAAWPAESHAPRAPSEDADSIYESVSEDGGRVYEEIPWVRVY  
 EN  
 25 ICLRRQDAGGAAPGDAPDSPYIEAENPLYDWGGSALFSPPGATRAPDPGLSLSEMPARPTNALAND  
 GP  
 TNVAALSALLTKLKRGRHQSH\*

30 Gene matched: gi|114350|sp|P10230|ATI2\_HSV11  
 Gene name: LPHA TRANS-INDUCING FACTOR

[SEQ ID NO:260]  
 35 ORF # = 50 from Contig 15  
 ORF start site = 93910  
 ORF end site = 92828  
 ORF sequence:  
 VGAAAVPLLSAGGAAPPHPGPDAAVFRSSLGSLLYWPGVRALLGRDCRVAARYAGRMTYIATGALLAR  
 40 FN  
 PGAVKCVLPREAAFAGRVLDVLAVLAEQTVQWLSVVVGARLHPHSAHPAFVDVEQEALFRALPLGSPG  
 VV  
 AAHEALGDTAARRLLATSGLNAVLGAAYALHTALATVTLKYALACGDARRRRDHAAAAAVLATGL  
 IL

QRLGLADTVVACVALAAFDGGSTAPEVGTYTPLRYACVLRATQPLYARTTPAKFWADVRAAAEHVDL  
RP  
ASSAPRAPVSGTADPAFLLEDLAAFPAPLNSESVLGPRVRVVDIMAQFRKLLMGDEETAALRAHVSG  
RR  
5 ATGLGGPPRP\*

Gene matched: gi|136920|sp|P10231|UL47\_HSV11

Gene name: VIRION PROTEIN UL47

10

[SEQ ID NO:261]

ORF # = 51 from Contig 15

ORF start site = 94919

15 ORF end site = 93504

ORF sequence:

MSVRGHAVRRRRRASTRSHAPSAHRADSPVEDEPEGGVGLMGYLRAVFNVDSEVEAAGEMASEEPP  
PR

20 RRREARGHPGSRRASEARAAAPRRASFPRPRSVTARSQSVRGRRDSAITRAPRGYLGPMDFRDVLG  
RV

GGSRVVPSPFLDELSEYEDDYPAAVAHDDGAGARPPATVEILAGRVSGPELQAAFPLDRLTFRVAAW  
DE

SVRSALALGHPAGFYPCPSAFGLSRVGMHFASPADPKVFFRQTLQQGEALAWYVTGDAILDLTDRR  
AK

25 TSPSRAMGFLVDAIVRVAINGWVCCTRLHTEGARLGARRQGGRAPTAVREPHGVAARGGRRRAAAQRG  
RG

RAPPPRRRRGLSQFAGVPAVLARGARAPGARLSRGRPLRGAHDVHRHRGSARPLQPRRRQMRAPAGG  
RV

CGARPGRAGGPGGADGPVALGGRGGAPAPALRPPRVCGRGAGGAVSRPAPG\*

30

Gene matched: gi|136920|sp|P10231|UL47\_HSV11

Gene name: VIRION PROTEIN UL47

35

[SEQ ID NO:262]

ORF # = 53 from Contig 15

ORF start site = 98257

ORF end site = 97349

40 ORF sequence:

MTSRRSVKSCPREAPRGTHEELYGVPSPADPESPRDDFRRGAGPMRARPRGEVRFLLHYDEAGYALYR  
DS

SSSEDNDESRTARPRRSASVAGSHGPGPARAPPPGGPVGAGGRSHAPPARTPKMTRGAPKAPATPA  
TD

PARGRRPAQADSAVLLDAPAPTASGRKTPAQGLAKKLHFSTAPPSPTAPWTPRVAGFNKRVFCAAVG  
RL  
AATHARLAAVQLWDMSRPHTDEDLNELLDLTTIRVTVCEGKNLLQRANELVNPDAAQDVDATAAARGR  
PA

5 GRAAATARAPARSASRPRRPLE\*

Gene matched: gi|136927|sp|P10233|UL49\_HSV11  
Gene name: TEGUMENT PROTEIN UL49

10

[SEQ ID NO:263]  
ORF # = 54 fro Contig 15  
ORF start site = 98876  
15 ORF end site = 98596  
ORF sequence:  
VVLLEFVALVAGVPGEPPNAAAGARGVIGDAQCRGDSAGVVSVPGLVLPFYLGMTSMGVCMIAHVYQICQ  
RA  
LAAGSA\*

20

Gene matched: gi|1944541|gnl|PID|e312365  
Gene name: (X14112) envelope protein [human]

25

[SEQ ID NO:264]  
ORF # = 55 from Contig 15  
ORF start site = 98867  
ORF end site = 99976  
30 ORF sequence:  
MSQWGPRAILVQTDSTNRNADGDWQAAVAIRGGGVQLNMVNKRAVDFTPAECGDSEWAVGRVSLGLR  
MA  
MPRDFCAIIHAPAVSGPGPHVMLGLVDSGYRGTVLAVVVAPNGTRGFAPGALRVDVTFLDIRATPPTL  
TE  
35 PSSLRHFPQLAPSPLAGLREDPWLDGALATAGGAVALPARRRGSLVYAGELTQVTTEHGDCVHEAPA  
FL  
PKREEDAGFDILIHRAVTVPANGATVIQPSLRVLRAADGPEACYVLGRSSLNARGLLVMPTRWPSGHA  
CA  
FVVCNLTGVPVTLQAGSKVAQLLVAGTHALPWIPPDNIHEDGAFRAYPRGVDPDATATPRDPPILVFTN  
40 EF  
DADAPPSKRGAGGFGSTGI\*

Gene matched: gi|118955|sp|P10234|DUT\_HSV11  
45 Gene name: DEOXYURIDINE 5'-TRIPHOSPHATE

- [SEQ ID NO:265]  
ORF # = 56 from Contig 15  
5 ORF start site = 101006  
ORF end site = 100182  
ORF sequence:  
MASLLGVLCGWGTRPEEQYEMIRAAAPPSXXDPRLQEALAVVNALLPAPITLDDALESDDTRRLVK  
AR  
10 ALARTYHACMVNLERLARHHPGLEGSTIDGAVAAHRDKMRRADTCMATILQMYMSVGAADKSADVLV  
SQ  
AIRSMAESDVVMEDVAIAERALGLSTSALAGGTRTAGLGATEAPPGPTRAQAPEVASVPVTHAGDRSP  
VR  
PGFVPPADPTPDRHRTSAPKRQASSTEAPLLLA\*  
15  
Gene matched: gi|136933|sp|P10235|UL51\_HSV1  
Gene name: PROTEIN UL51
- 20 [SEQ ID NO:266]  
ORF # = 58 from Contig 15  
ORF start site = 102815  
ORF end site = 104188  
ORF sequence:  
25 MYVNRNEIFNAALAVTNIILDLIDIALKEPVFPFRLHEALGHFRRGALAAVQLLFPAARVDPDAYPCYF  
FK  
SACRPRAPPVCAGDGPSAGGDDGDWFPDAGGDDGDEWEEDTDPMDTTHGPLPDDEAAYLDLLHEQ  
IP  
AATPSEPDSVVCSCADKIGLRVCLPVPAPYVVHGS�TMRGVARVIQQAVLLDRDFVEAVGSHVKNFLL  
30 ID  
TGVYAHGHSRLRPYFAKIGPDGSACGRLLPVFVIPPACEDVPAFVAHADPRRFHFHAPPMFSAAPRE  
IR  
VLHSLGGDYVSFFEKKASRNALEHFGRRETLTEVLGRYDVDPDAGETVEGFASELLGRIVACIEAHFP  
EH  
35 AREYQAVSVRRRAVIKDDWLLQLIPGRGALNQSLSCLRFKHGRASRATARTFLALSVGTNNRLCASLC  
QQ  
CFATKCDNNRLHTLFTVDAGTPCSRSAPSSTSRSPPSSS\*
- 40 Gene matched: gi|136939|sp|P10236|UL52\_HSV11  
Gene name: DNA HELICASE/PRIMASE COMPLEX
- [SEQ ID NO:267]  
45 ORF # = 59 from Contig 15

ORF start site = 104140  
ORF end site = 105156  
ORF sequence:  
MLAVRSLQHLLTVIFITAYGLVLAWYIVFGASPLHRCIYAVRPAGAHNDTALVWMKINQTLFLGPPT  
5 AP  
PGGAWTPHAHVCYANIIEGRAVSLPAIPGAMSRVMNVHEAVNCLEALWDTQMRLVVVGWFLYLAFVA  
LH  
QRRCMFGVSPAHSMVAPATYLLNYAGRIVSSVFLQYPYTKITRLLCELSVQRQTLVQLFEADPVTFL  
YH  
10 RPAVGVI VGCCELLLRFVALGLIVGTALISRGACAITYPEFLTITTTWCFVSIIALTELYFILRRDSAPK  
NA  
EPAAPRGRSKGWSGVCGRCCSIILSGIAVRLCYIAVVAGVVLALRYEQEIQRRLFDL\*

15 Gene matched: gi|116105|sp|P22485|CELF\_HSV2H  
Gene name: CELL FUSION PROTEIN PRECURSO

[SEQ ID NO:268]  
20 ORF # = 60 from Contig 15  
ORF start site = 105702  
ORF end site = 107240  
ORF sequence:  
MATDIDMLIDLGLDLSDELEEDALERDEEGRRDDPESDSSGECSSSDEDMEDPCGDGGAEIDAIAIP  
25 KG  
PPARPEDAGTPEASTPRPAARRGADDPPTTGVSRLGTRRSASPREPHGKGVARIQPPSTKAPHPR  
GG  
RRGRRRGRGRYGPGGADSTPNPRRRVSRNAHNQGRHPASARTDGP GATHGEARRGGEQLDVSGGPRP  
RG  
30 TRQAPPPLMALSLTPPHADGRAPVPERKAPSADTIDPAVRAVLRSISERAVERISESFGRSALVMQD  
PF  
GGMPFFPAANSPWAPVLATQAGGFDAETRRVSWETLVAHGPSLYRTFAANPRAASTAKAMRDCVLRQEN  
LI  
EALASADETLAWCKMCIHHNLPLRPQDPIIGTAAAVLENLATRLRPFLQCYLKARGLCGLDDLCSRRR  
35 LS  
DIKDIAFVLVILARLANRVERGVSEIDYTTVGVGAGETMHFYIPGACMAGLIEILDTHRQECSSRVC  
EL  
TASHTIAPLYVHGKYFYCNSLF\*

40 Gene matched: gi|124181|sp|P28276|IE63\_HSV2H  
Gene name: TRANSCRIPTIONAL REGULATOR IE

ORF # = 62  
45 ORF start site = 108542



ORF end site = 108189  
 ORF sequence:  
 MIGAHPGVGGDLPSGLPTYAEATSDRPPTYAMVMAACPTEPPGGSVGPADQPRVQSSRTWRPPLVNSR  
 EL  
 5 YRAQRAARCASSSDTPQAPGWCGGTCRHAVFGVVAVVVVVIILAF LWR\*

Gene matched: gi|136952|sp|P28282|UL56\_HSV2H  
 Gene name: PROTEIN UL56•gi|73833|pir||W  
 10

[SEQ ID NO:269]  
 ORF # = 63 from Contig 15  
 ORF start site = 112958  
 15 ORF end site = 113542  
 ORF sequence:  
 MHLFCQCPLTDGQDLYLCPVYPRMHQEHLVCPLHRLDDARRRGRTSAAWDEGLVRALTHSGGLMGCGG  
 RS  
 LTLSETYWGHPLEYEKLVPWDHPRDLKVPEASAVGTRALVPRGRGRPLRGRPVPLIPLDCEPNDGLPFG  
 20 GG  
 WPGGRLRGAPVPLHPPPPSAPPLSFTPTLTTPCLCRGLSLCVVVKQYLKDRNNF\*

Gene matched: gi|1644457  
 25 Gene name: (U72521) neural variant mena+ protein [Mus muscu

30

#### TABLE 4

35 All amino acid sequences within Table 4 are encoded by Contig 15  
 [SEQ ID NO:217] of Table 3

[SEQ ID NO:270] = 15  
 ORF # = 9b  
 40 ORF start site = 14508  
 ORF end site = 11905  
 ORF sequence:  
 MNVATCTHQTHHAARAPGATSAPGAASGDPLGARRPIGDDECEQYTSSVSLARMLYGGDLAEWVPRVH  
 PK

TTIERQQHGPVTFPDASAPTARCVTVVRAPMSGGKTTALIRWLGEAIHSPDTSVLVVSCRRSFTQTLA  
TR  
FAESGLPDFVTYFSSTNYIMNDRPFHRLIVQVESLHRVGNLLNNYDVLVLDEVMSTLGQLYSPTMQQ  
LG  
5 RVDALMLRLLRTPRIIAMDATANAQLVDFLC SLRGEKNVHV VIG EYAMP GFSARRCLFLPRLGPEVL  
QA  
ALRPPGPAGGAPPPDAPPDATFFGELEARLAGGDNVCIFSSTVSFAEVVARFCRQFTDRVLLHSLTP  
PG  
DVTTWGRYRVVIYTTVVTVGLSFDPPHFDSMFAYVKPMNYGPDMSVYQSLGRVRTLRKGELLIYMDG  
10 SG  
ARSEPVTTPMLLNHVVSASGQWPAQFSQVTNLLCRRFKGRCDASHADAAQARGSRIYSKFRYKHYFER  
CT  
LACLADSLNILHMLLTLCNMHVRFWGHDAALTPRNFCLFLRGIHFDALRAQRDLRELRCQDPDTSLSA  
QA  
15 AETEEVGLFVEKYLRPDVAPAEVVALMRGLNSLVGRTRFIYLVLEACLRVPMAAHSSAIFRRLYDHY  
AT  
GVIPTINAAGELELVALHPTLN VAPVWELFRLCSTMAACLQWDSMAGGSGRTFSPEDVLELLNPHYDR  
YM  
QLV FELGHCNVT DGPLLED AVKR VADALSGC PPRGSVSETEHALSLFKIIWGELFGVQLAKSTQTFP  
20 GA  
GRVKNLTKRAIVELDDAHRIDHSACRTHRQLYALLMAHKREFAGARFKLRAPAWGRCLRTHASGAQPN  
TD  
IILEAALSELPT EAWPM MQAVNFSTL\*

25  
Gene matched: gi|1869831|gnl|PID|e304265  
Gene name: (286099) UL9 [human herpesvirus

[SEQ ID NO:271] = 15  
30 ORF # = 9a  
ORF start site = 14520  
ORF end site = 11904  
ORF sequence:  
MAETMNVATCTHQTHHAARAPGATSAPGAASGDPLGARRPIGDDECEQYTSSVSLARMLYGGDLAEWV  
35 PR  
VHPKTTIERQQHGPVTFPDASAPTARCVTVVRAPMSGGKTTALIRWLGEAIHSPDTSVLVVSCRRSFT  
QT  
LATRFAESGLPDFVTYFSSTNYIMNDRPFHRLIVQVESLHRVGNLLNNYDVLVLDEVMSTLGQLYSP  
TM  
40 QQLGRVDALMLRLLRTPRIIAMDATANAQLVDFLC SLRGEKNVHV VIG EYAMP GFSARRCLFLPRLG  
PE  
VLQAALRPPGPAGGAPPPDAPPDATFFGELEARLAGGDNVCIFSSTVSFAEVVARFCRQFTDRVLLH  
SL  
TPPGDVTTWGRYRVVIYTTVVTVGLSFDPPHFDSMFAYVKPMNYGPDMSVYQSLGRVRTLRKGELLI  
45 YM

DGSGARSEPVFTPMLLNHVVSASGQWPAQFSQVTNLLCRRFKGRCDASHADAAQARGSRIYSKFRYKH  
 YF  
 ERCTLAACLADSLNHLMLLTLCNMHVRFWGHDAALTPRNFCLFLRGIHFDALRAQRDLRELRCQDPDT  
 SL  
 5 SAQAAETEEVGLFVEKYLRPDVAPAEVVALMRGLNSLVGRTRFIYLVLEACLRVPMAAHSSAIFRRL  
 YD  
 HYATGVIPTINAAGELELVALHPTLNVA PVWELFRLCSTMAACLQWDSMAGGSGRTFSPEDVLELLNP  
 HY  
 DRYMQLVFELGHCNVTDGPLLSEDAVKRVADALSGCPPRGSVSETEHALSLFKIIWELFGVQLAKST  
 10 QT  
 FPGAGRVKNLTKRAIVELLDAHRIDHSACRTHRQLYALLMAHKREFAGARFKLRAPAWGRCLRTHASG  
 AQ  
 PNTDIILEAALSELPTEAWPMMQGA VNFSTL\*  
 15  
 Gene matched: gi|136806|sp|P10193|OBP\_HSV11  
 Gene name: ORIGIN OF REPLICATION BINDING  
 20 [SEQ ID NO:272] = 15  
 ORF # = 20a  
 ORF start site = 31782  
 ORF end site = 27630  
 ORF sequence:  
 25 MEPANPPRNPMAPARDPPGYRYAAAMVPTGSILSTIEVASHRRLFDF FARVRSDENSLYDVEFDALL  
 GS  
 YCNTLSLVRFLLEGLSVACVCTKFPELAYMNEGRVQFEVHQPLIARDGPHPVVEQPVHNYMTKVIDRRA  
 LN  
 AAFSLATEAIALLTGEALDGTGISLHRQLRAIQQLARNVQAVLGAFERGTDQMLHVLLEKAPPLALL  
 30 LP  
 MQRYL DNGRLATRVARATLVAELKRSFCDTSFFLGKAGHRREAIEAWLVDLTATQPSVAVPRLTHAD  
 TR  
 GRPVDGVLVTAAIKQRLQSF LKVEDTEADVPTYGEMVLNGANLVTALVMGKAVRSLDDVGRHLE  
 MQ  
 35 EEQLEANRETLELESAPQTTRVRADLVAIGDRLVFLEALEKRIYAATNPYPYPLVGAMDLTFVLPLGL  
 FN  
 PAMERFAAHAGDLVPAPGHPEPRAFP PPRQLFFWGKD HQVLR L SMENAVGT VCHPSLMNIDA AVGGVNH  
 DP  
 VEAANPYGAYVAAPAGPGADMQQRFLNAWRQRLAHGRVVRWVAECQM TAEQFMQPDNANLAL E LHPAFD  
 40 FF  
 AGVADVELPGGEVPPAGPGAIQATWRV VNGNLPLALCPVAFRDARGLELGVGRHAMAPATIAAVRGAF  
 ED  
 RSYPAVFYLLQAAIHGSEHVFCALARLV TQCITSYWNNT RCAA FVNDYSLVSYIVTYLGGDLPEECMA  
 VY

RDLVAHVEALAQLVDDFTLPGPELGGQAELNHLMRDPALLPPLVWDCDGLMRHAALDRHRDCRIDA  
 GG  
 HEPVYAAACNVATADFNRNDGRLLHNTQARAADAADDRPHRPADWTVHHKIYYYVLVPAFSRGRCTA  
 GV  
 5 RFDRVYATLQNMVVPEIAPGEECPSPVTPDAHPLHPANLVANTVNAMFHNGRVVVDGPAMLTQVLA  
 HN  
 MAERTTALLCSAAPDAGANTASTANMRIFDGALHAGVLLMAPQHLDHTIQNGEYFYVLPVHALFAGAD  
 HV  
 ANAPNFFPALRDLARHVPLVPPALGANYFSSIRQPVVQHARESAAGENALTYALMAGYFKMSPVALYH  
 10 QL  
 KTGLHPGFGFTVVRQDRFVTENVLFSERASEAYFLGQLQVARHETGGGVSTLTQPRGNVDLGVGYTA  
 VA  
 ATATVRNPVTDMGNLQNFYLGARGAPPLLDNAAVYL RNAVVAGNRLGPAQPLPVFGCAQVPRRAGMD  
 HG  
 15 QDAVCEFIATPVATDINYFRPCNPRGRAAGGVYAGDKEGDVIALMYDHGQSDPARPFAATANPWASQ  
 RF  
 SYGDLLYNGAYHLNGASPVLSPCFKFFTAADITAKHRCLERLIVETGSAVSTATAASDVQFKRPPGCR  
 EL  
 VEDPCGLFQEAYPITCASDPALLRSARDGEAHARETHFTQYLIYDASPLKGLSL\*

20  
 Gene matched: gi|137571|sp|P06491|VCAP\_HSV11  
 Gene name: MAJOR CAPSID PROTEIN (MCP)

25 [SEQ ID NO:273] = 15  
 ORF # = 20b  
 ORF start site = 31754  
 ORF end site = 27630  
 ORF sequence:

30 MAAPARDPPGYRYAAAMVPTGSILSTIEVASHRRLFDFARVRSDENSLYDVEFDALLGSYCNLTSLV  
 RF  
 LELGLSVACVCTKFPDELAYMNEGRVQFEVHQPLIARDGPHPVEQPVHNYMTKVIDRRALNAAFSLATE  
 AI  
 ALLTGEALDGTGISLHRQLRAIQQLARNVQAVLGAFERGTADQMLHVLLKAPPLALLLPMQRYLDNG

35 RL  
 ATRVARATLVAELKRSFCDTSFFLGKAGHRREAIEAWLVDLTTATQPSVAVPRLTHADTRGRPVDGVL  
 VT  
 TAAIKQRLQLQSFLKVEDTEADVPTYGEMVLNGANLVTALVMGKA VRSLDDVGRHLLMQEEQLEANR  
 ET

40 LDELESAPQTTRVRADLVAIGDRLVFLEALEKRIYAATNPYPYPLVGAMDLTFLVPLGLFNPAMERFAA  
 HA  
 GDLVPAPGHPEPRAFPQRQLFFWGKDHQVLRLSMENAVGTVCPSLMNIDAAVGGVNHDPVEAANPYG  
 AY  
 VAAPAGPGADMQQRFLNAWRQRLAHGRVRVWAECQMTAEQFMQPDNANLALHLPADFDFAGVADVEL

45 PG

GEVPPAGPGAIQATWRVVNGNLPLALCPVAFRDARGLELGVGRHAMAPATIAAVRGAFEDRSYPVAFY  
 LL  
 QAAIHGSEHVFCALARLVTQCITSYWNTRCAAFVNDYSLVSYIVTYLGGDLPEECMAVYRDLVAHVE  
 AL  
 5 AQLVDDFTLPGPELGGQAQAEHLNMRDPALLPPLVWDCDGLMRHAALDRHRDCRIDAGGHEPVYAAA  
 CN  
 VATADFNRRNDGRLHNTQARAADAADDRPHRPADWTVHHKIYYYVLVPAFSRGRCCCTAGVRFDRVYAT  
 LQ  
 NMVVPEIAPGEECPSPDPTDPAHPLHPANLVANTVNAMFHNGRVVVDGPAMLTQLVLAHNMAERTTAL  
 10 LC  
 SAAPDAGANTASTANMRIFDGALHAGVLLMAPQHLDTIQNGEYFYVLPVHALFAGADHVNANPNFPP  
 AL  
 RDLARHVPLVPPALGANYFSSIRQPVVQHARESAAGENALTYALMAGYFKMSPVALYHQLKTGLHPGF  
 GF  
 15 TVVRQDRFVTENVLFSERASEAYFLGQLQVARHETGGVSFTLTQPRGNVDLGVGYTAVAATATVRNP  
 VT  
 DMGNLPQNFYLGRGAPPLLDNAAVYLRNAVAVAGNRLGPAQPLPVFGCAQVPRRAGMDHGQDAVCEFI  
 AT  
 PVATDINYFRRPCNPRGRAAGGVYAGDKEGDVIALMYDHGQSDPARPFAATANPWASQRFSGDLLYN  
 20 GA  
 YHLNGASPVLSPCFKFFTAADITAKHRCLERLIVETGSAVSTATAASDVQFKRPPGCRELVEDPCGLF  
 QE  
 AYPITCASDPALLRSARDGEAHARETHFTQYLIYDASPLKGLSL\*  
 25 Gene matched: gi|137571|sp|P06491|VCAP\_HSV11  
 Gene name: MAJOR CAPSID PROTEIN (MCP)

[SEQ ID NO:274] = 15  
 30 ORF # = 22a  
 ORF start site = 33002  
 ORF end site = 34984  
 ORF sequence:  
 MRPELSLKGRPCVTEAVVCPSTDAAIHSGGSSSVRPQPYARAARARATHGSRSRHRQPLPPSSSHHP  
 35 TI  
 PPPSPPRGSPAMELSYATTLHHRDVVFYVTADRNRAYFVCGGSVYSVGRPRDSQPGEIAKFGLVVRG  
 TG  
 PKDRMVANYVRSELQRGLRDVRPVGEDEVFLDSVCLLNPNVSSERDVINTNDVEVLDECLAEYCTSL  
 RT  
 40 SPGVLVTGVRVRARDRVIELFEHPAIVNISSRFAYTPSPYVFALQAHLPRLPSSLEPLVSGLFDGIP  
 AP  
 RQPLDARDRTDVVITGTRAPRPMAGTGAGGAGAKRATVSEFVQVKHIDRVVSPSVSSAPPPSAPDAS  
 LP  
 PPGLQEAAPPGPPLRELWWVFYAGDRALEEPHAESGLTREEVRAVHGFREQAWKLFSGVAPRAFLGA  
 45 AL

ALSPTQKLAVYYYLIHRERRMSPPALVRLVGRYIQRHGLYVPAPDEPTLADAMNGLFRDALAAGTVA  
EQ  
LLMFDLLPPKDVPGSDARADSAALLRFVDSQRLTPGGSVSPHEVMYLGAFLGVLYAGHGRLAAATH  
AR

5 LTGVTSLVLTVGDDVRMSAFDRGPAGAAGRTRTAGYLDALLTVCLARAQHGQSV\*

Gene matched: gi|136845|sp|P10205|UL21\_HSV11

Gene name: PROTEIN UL21

10

[SEQ ID NO:275] = 15

ORF # = 22b

ORF start site = 33385

ORF end site = 34984

15 ORF sequence:

MELSYATTLHHRDVVFYVTADRNRAYFVCGGSVSVGRPRDSQPGEIAKFGLVVRGTGPKDRMVANYV  
RS  
ELRQRLRDVRPVGEDVFLDSVCLLNPNVSSERDVINTNDVEVLDECLAEYCTSLRTSPGVLVTGVR  
VR

20 ARDRVIELFEHPAIVNISSRFAYTPSPYVFALAQAHLPRLPSSLEPLVSGLFDGIPAPRQPLDARDRR  
TD

VVITGTRAPRPMAGTGAGGAGAKRATVSEFVQVKHIDRVVSPSVSSAPPPSAPDASLPPGLQEAAPP  
GP

PLRELWWVFYAGDRALEEPHAESGLTREEVRAVHGFREQAWKLFSGVGAAPRAFLGAALALSPTQKLAV

25 YY

YLIHRERRMSPPALVRLVGRYIQRHGLYVPAPDEPTLADAMNGLFRDALAAGTVAEQLLMFDLLPPK  
DV  
PVGSDARADSAALLRFVDSQRLTPGGSVSPHEVMYLGAFLGVLYAGHGRLAAATHARLTGVTSLVLT  
VG

30 DVDRMSAFDRGPAGAAGRTRTAGYLDALLTVCLARAQHGQSV\*

Gene matched: gi|136845|sp|P10205|UL21\_HSV11

Gene name: PROTEIN UL21

35

[SEQ ID NO:276] = 15

ORF # = 41a

ORF start site = 75756

ORF end site = 77588

40 ORF sequence:

MTAAALYGGAKYRPGTLRNPGRVASTPRRRGVLYGALCPGIPFVGSGPGAVGWECVCVGGGRRDGGPD  
QV  
YRGRSVGRPNRPFKHLRMHRPSQSDTGTHQRRKPPSPVRVRVFSGGVFFLSALLPPHLHHPPTTRPL  
AI

- GGKTMKTKPLPTAPMAWAESAVETTTSPRELAGHAPLRRVLRPPIARRDGPVLLGDRAPRRTASTMWL  
 LG  
 IDPAESSPGTRATRDDTEQAVDKILRGARRAGGLTVPGAPRYHLTRQVTLTDLCPNAERAGALLLAL  
 RH
- 5 PTDLPHLARHRAPPGRQTERLAEAWGQLEASALGSGRAESGCARAGLVSFNFLVAACAAAYDARDAA  
 EA  
 VRAHITNTNYGGTRAGARLDRFSECLRAMVHTHVFPHEVMRFFGGLVSWVTQDELASVTAVCSGPQEAT  
 HT  
 GHPGRPCSAVTIPACAFVDLDAELCLGGPGAFLYLFTYRQCRDQELCCVYVKSQLPFRGLEAALE
- 10 RL  
 FGRLRITNTIHGAEDMTPPPNRNVDPLAVLAASSQSPRCSASQVTNPQFVDRLYRWQPDLRGRPTA  
 RT  
 CTYAFAELGVMPDNPRLHRTERFGAVGVPVILEGVVWRPVGWRACA\*
- 15 Gene matched: gi|139176|sp|P22486|VP19\_HSV2G  
 Gene name: CAPSID ASSEMBLY AND DNA MATU
- [SEQ ID NO:277] = 15
- 20 ORF # = 41b  
 ORF start site = 75817  
 ORF end site = 77588  
 ORF sequence:  
 MHRPSQSDTGTHQRRKPPSPVRVRVFSGGVFFLSALLPPLHHPPTTRPLAIGKTMKTKPLPTAPM
- 25 AW  
 AESAVETTTSPRELAGHAPLRRVLRPPIARRDGPVLLGDRAPRRTASTMWLLGIDPAESSPGTRATRD  
 DT  
 EQAVDKILRGARRAGGLTVPGAPRYHLTRQVTLTDLCPNAERAGALLLALRHPTDLPHLARHRAPP  
 RQ
- 30 TERLAEAWGQLEASALGSGRAESGCARAGLVSFNFLVAACAAAYDARDAEAVRAHITNTNYGGTRAG  
 AR  
 LDRFSECLRAMVHTHVFPHEVMRFFGGLVSWVTQDELASVTAVCSGPQEATHTGHPGRPCSAVTIPAC  
 AF  
 VDLDAELCLGGPGAFLYLFTYRQCRDQELCCVYVKSQLPFRGLEAALERLFGRLRITNTIHGAED
- 35 MT  
 PPPPNRNVDPLAVLAASSQSPRCSASQVTNPQFVDRLYRWQPDLRGRPTARTCTYAFAELGVMPDN  
 SP  
 RCLHRTERFGAVGVPVILEGVVWRPVGWRACA\*
- 40 Gene matched: gi|139176|sp|P22486|VP19\_HSV2G  
 Gene name: CAPSID ASSEMBLY AND DNA MATU
- [SEQ ID NO:278] = 15
- 45 ORF # = 41c

ORF start site = 76188

ORF end site = 77588

ORF sequence:

MKTKPLPTAPMAWAESAVETTTSPRELAGHAPLRRVLRPPIARRDGPVLLGDRAPRRTASTMWLLGID  
5 PA  
ESSPGTRATRDDTEQAVDKILRGARRAGGLTVPGAPRYHLTRQVTLTDLCPNAERAGALLALRHPT  
DL  
PHLARHRAPPGRQTERLAEAWGQLLEASALGSGRAESGCARAGLVSFNFLVAACAAAYDARDAEAVR  
AH  
10 ITTNYGGTRAGARLDRFSECLRAMVHTHVFPHEVMRFFGGGLVSWVTQDELASVTAVCSGPQEATHHTGH  
PG  
RPCSAVTIPACAFVDLDAELCLGGPGAFLYLVTYRQCRDQELCCVYVVKSQLPPRGLEAALERLFG  
RL  
RITNTIHGAEDMTPPPPNRNVDFPLAVLAASSQSPRCSASQVTNPQFVDRLYRWQPDLRGRPTARTCT  
15 YA  
AFAELGVMPDNSPRCLHRTERFGAVGVPVVILEGVVWRPGGWRACA\*

Gene matched: gi|139176|sp|P22486|VP19\_HSV2G

Gene name: CAPSID ASSEMBLY AND DNA MATU

20

[SEQ ID NO:279] = 15

ORF # = 46a

ORF start site = 86432

25 ORF end site = 87820

ORF sequence:

MAVVCSGSLRLRPFHPPSPSFFVLRALIRAGPGPFAASPRAPSGPGCGMCRGDSPGVAGGSGEHCLGG  
DD  
GDDGRPRLACVGAIARGFAHLWLQATTLGFGVSVVLSRGPYADAMSGAFVIGSTGLGFLRAPPAFARP  
30 PT  
RVCAWLRLVGGGAVALWSLGEAGPPGVPGPATQCLALGAAYAALLVLADDVHPLFLLAPRPLFVGT  
LG  
VVVGGLTIGGSARYWWIDPRAAAALTAAVVAGLGTAAAGDSFSKACPRHRRFCVVS AVESPPPRYAPE  
DA  
35 ERPTDHGPLL PSTHHQRS PRVCGDGAARPENIWVPVVT FAGALALAACAARGSDAAPSGPVLPLWPQV  
FV  
GGHAAAGLTELCTLAPRDLTDPLL FAYVGFQV VNHGLMFVVPDIAVYAMLGGAVWISLTQVLGLRRR  
LH  
KDPDAGPWAAATLRGLFFSVYALGFAAGVLVRPRMAASRRSG\*

40

Gene matched: gi|136909|sp|P10227|UL43\_HSV11

Gene name: MEMBRANE PROTEIN UL43

45 [SEQ ID NO:280] = 15



ORF # = 46b

ORF start site = 86576

ORF end site = 87820

ORF sequence:

5 MCRGDSPGVAGGSGEHCLGGDDGDDGRPRLACVGAIARGFAHLWLQATTLGFGSVVLSRGPYADAMS  
GA  
FVIGSTGLGFLRAPPAFARPPTRVCAWLRLVGGGAVALWSLGEAGAPPGVPGPATQCLALGAAYAAL  
LV  
LADDVHPLFLLAPRPLFVGTGLGVVGGTLIGGSARYWWIDPRAAAALTAAVVAGLGTAAAGDSFSKAC  
10 PR  
HRRFCVVS AVESPPPRYAPEDAERPTDHGPLL PSTHHQRS PRVCGDGAARPENIWVPVVT FAGALALA  
AC  
AARGSDAAPSGPVLPLWPQVFVGGHAAAGLTEL CQTLAPRDLTDPLL FAYVGFQVNVHGLMFVVPDIA  
VY  
15 AMLGGAVWISLTQVLGLRRRLHKDPDAGPWAAATLRGLFFSVYALGFAAGVLVRPRMAASRRSG\*

Gene matched: gi|136909|sp|P10227|UL43\_HSV11

Gene name: MEMBRANE PROTEIN UL43

20

[SEQ ID NO:281] = 15

ORF # = 57a

ORF start site = 100984

ORF end site = 102942

25 ORF sequence:

MGTEDCDHEGRSVAAPVEVMALYATDGCVITSS LALLTNCLLGA EPLYIFS YDAYRPDAPNGPTGAPT  
EQ  
ERFEGSRALYRDAGGLNGDSFRVTFCLLGTEVGVT HHPKGRTRPMFVCRFERADDVA VLQDALGRGTP  
LL  
30 PAHITATLDLEATFALHANIIMALTVAIVHNAPARIGSGSTAPLYEPGESMR SVVGRMSLGQRGLTTL  
FV  
HHEARVLAAYRRAYY GSAQSPFWFLSKFGPDEKSLVLAARYYVLQAPRLGGAGATYDLQAVKD ICATY  
AI  
PHDPRPDTLSAASLT SFAAITRFCCTSQYSRGAAAAGFPLYVERRIAADVRETGALEKFI AHDRSCLR  
35 VS  
DREFITYIYLAHFE CFSPRLATHLRAVTTHDPSPAASTE QPSPLGREAVEQFFRHVRAQLNIREYVK  
QN  
VTPRETALAGDAAAAYLRARTYAPAALTPAPAYCGVADSSTKMMGR LAEAERLLVPHGWPAFAPTT PG  
DD  
40 AGGGTAAPQTCGIVKRL LKLAATEQQGTTPPAIAALMQDASVQTPLPVYRITMSPTGQAF AAAARD DW  
AR  
VTRDARPP EATVVADAAA APEPGALGRRLTRRICARGPAPPPGRPGRRGPDVREPQRDLQRRAGRYEH  
HP  
GSGHRPEGARPLSPAPRGP GSL\*

45

Gene matched: gi|136939|sp|P10236|UL52\_HSV11  
 Gene name: DNA HELICASE/PRIMASE COMPLEX

5 [SEQ ID NO:282] = 15  
 ORF # = 57b  
 ORF start site = 100765  
 ORF end site = 102942  
 ORF sequence:  
 10 MGAGKSALTARASCSRGXSSEGGAAARIISYCCSSGRVPQPHSTPSRDAIPEHARGSAPAFPHPTPS  
 GF  
 AGAMGTEDCDHEGRSVAAPVEVMALYATDGCVITSSSLALLTNCLLGAEPLYIFSYPDAYRPDAPNGPTG  
 AP  
 TEQERFEGSRALYRDAGGLNGDSFRVTFCLLGTEVGVTTHHPKGRTRPMFVCRFERADDDAVLQDALGR  
 15 GT  
 PLLPAHITATLDLEATFALHANIIMALTVAIVHNAPARIGSGSTAPLYEPGESMRSVVGRMSLGQRGL  
 TT  
 LFBVHHEARVLAAYRRAYYGSAQSPFWFLSKFGPDEKSLVLAARYVVLQAPRLGGAGATYDLQAVKDIC  
 AT  
 20 YAIPHDPRPDTLAASLTSFAAITRFCCTSQYSRGAAGFPLYVERRIAADVRETGALEKFIHSDRS  
 CL  
 RVSDREFITYIYLAHFECFSPRLATHLRAVTTHDPSPAASTEQPSPLGREAVEQFFRHVRAQLNIRE  
 YV  
 KQNVTPRETALAGDAAAAYLRARTYAPAALTPAPAYCGVADSSTKMMGRLEAEERLLVPHGWPAFAPT  
 25 TP  
 GDDAGGGTAAPQTCGIVKRLKLAATEQQGTTTPAIAALMQDASVQTPLPVYRITMSPTGQAFAAAAA  
 DD  
 WARVTRDARPEATVVADAAAPEPGALGRRLTRRICARGPAPPPGRPGRRGPDVREPQORDLQRRAGR  
 YE  
 30 HHPGSGHRPEGARPLSPAPRGPGL\*

Gene matched: gi|136939|sp|P10236|UL52\_HSV11  
 Gene name: DNA HELICASE/PRIMASE COMPLEX

35 [SEQ ID NO:283] = 15  
 ORF # = 57c  
 ORF start site = 100678  
 ORF end site = 102942  
 40 ORF sequence:  
 MQAWYVRARARAFTRRRVSSSDSRASSVVMGAGKSALTARASCSRGXSSEGGAAARIISYCCSSGRV  
 PQ  
 PHSTPSRDAIPEHARGSAPAFPHPTPSGFAGAMGTEDCDHEGRSVAAPVEVMALYATDGCVITSSSLA  
 LT

NCLLGAEPLYIFSYDAYRPDAPNGPTGAPTEQERFEGSRALYRDAGGLNGDSFRVTFCLLGTEVGVTH  
 HP  
 KGRTRPMFVCRFERADDDAVLQDALGRGTPLLPAHITATLDLEATFALHANIIMALTVAIVHNAPARI  
 GS  
 5 GSTAPLYEPGESMRSVVGRMSLGQRGLTTLFVHHEARVLAAYRRAYYGSAQSPFWFLSKFGPDEKSLV  
 LA  
 ARYYVLQAPRLGGAGATYDLQAVKDICATYAIPHDPRPDTLSAASLTSFAAITRFCCTSQYSRGA  
 GF  
 PLYVERRIAADVRETGALEKFIHADRSLRVSDREFITYIYLAHFECFSPRLATHLRAVTTHDPSA  
 10 AS  
 TTEQPSPLGREAVEQFFRHRAQLNIREYVKQNVTPRETALAGDAAAAYLRARTYAPAALTPAPAYCGV  
 AD  
 SSTKMMGRLEAEERLLVPHGWPAFAPTPGDDAGGGTAAPQTCGIVKRLCLKLAATEQQGTPPAIAAL  
 MQ  
 15 DASVQTPLPVYRITMSPTGQAFAAAARDWARVTRDARPPPEATVADAAAAPEPGALGRRLTRICAR  
 GP  
 APPPGRPGRRGPDVREPQRDQLRRAGRYEHHPGSGHRPEGARPLSPAPRGPGSL\*

Gene matched: gi|136939|sp|P10236|UL52\_HSV1

20 Gene name: DNA HELICASE/PRIMASE COMPLEX

[SEQ ID NO:284] = 15  
 ORF # = 57d  
 25 ORF start site = 100624  
 ORF end site = 102942  
 ORF sequence:  
 MVEPSSPGWWRASLSRLTMQAWYVRARARAFTRRRVSSSDSRASSSVMGAGKSALTARASCSRGXS  
 EG  
 30 GAAARIISYCCSSGRVPQPHSTPSRDAIPEHARGSAPAFPHPTPSGFAGAMGTEDCDHEGRSVAAPVE  
 VM  
 ALYATDGCVITSSLALLTNCLLGAEPLYIFSYDAYRPDAPNGPTGAPTEQERFEGSRALYRDAGGLN  
 DS  
 FRVTFCLLGTEVGVTHHPKGRTRPMFVCRFERADDDAVLQDALGRGTPLLPAHITATLDLEATFALHA  
 35 NI  
 IMALTVAIVHNAPARIGSGSTAPLYEPGESMRSVVGRMSLGQRGLTTLFVHHEARVLAAYRRAYYGSA  
 QS  
 PFWFLSKFGPDEKSLVLAARYYVLQAPRLGGAGATYDLQAVKDICATYAIPHDPRPDTLSAASLTSFA  
 AI  
 40 TRFCCTSQYSRGAAGGFPLYVERRIAADVRETGALEKFIHADRSLRVSDREFITYIYLAHFECFSP  
 PR  
 LATHLRAVTTHDPSAASTEQPSPLGREAVEQFFRHVRAQLNIREYVKQNVTPRETALAGDAAAAYLR  
 AR  
 TYAPAALTPAPAYCGVADSSTKMMGRLEAEERLLVPHGWPAFAPTPGDDAGGGTAAPQTCGIVKRL  
 45 KL

AATEQQGTTTPAIAALMQDASVQTPLPVYRITMSPTGQAFAAAARDWARVTRDARPEATVVADAAA  
AP  
EPGALGRRLTRRICARGPAPPPGRPGRRGPDVREPQDRLQRRAGRYEHHPGSGHRPEGARPLSPAPRG  
PG

5 SL\*

Gene matched: gi|136939|sp|P10236|UL52\_HSV11  
Gene name: DNA HELICASE/PRIMASE COMPLEX

10

[SEQ ID NO:285] = 15  
ORF # = 57e  
ORF start site = 100567  
RF end site = 102942

15

ORF sequence:  
MHVSARRRILSRCAATAPSMVEPSSPGWWRASLSRLTMQAWYVRARARAFTRRRVSSSDSRASSVMG  
AG  
KSALTARASCSRGXSSEGGAARIISYCCSSGRVPQPHSTPSRDAIPEHARGSAPAFPHPTPSGFAG  
AM

20

GTEDCDHEGRSVAAPVEVMALYATDGCVITSSLLLTNCLLGAEPYIFSYDAYRPDAPNGPTGAPTE  
QE  
RFEGRALYRDAGGLNGDSFRVTFCLLGTEVGVTTHPKGRTRPMFVCRFERADDVAVLQDALGRGTPL  
LP

25

AHITATLDLEATFALHANIIMALTVAIVHNAPARIGSGSTAPLYEPGESMRVSVGRMSLGQRGLTTLF  
VH  
HEARVLAAYRRAYYGSQSPFWFLSKFGPDEKSLVLAARYVVLQAPRLGGAGATYDLQAVKDICATYA  
IP  
HDPRPDTLSAASLTSFAAITRFCCTSQYSRGAAGFPLYVERRIAADVRETGALEKFAHDRSCLRV  
SD

30

REFITYIYLAHFECFSPRLATHLRAVTTHDPSPAASTEQPSPLGREAVEQFFRHVRAQLNIREYVKQ  
NV  
TPRETALAGDAAAAYLRARTYAPAALTPAPAYCGVADSSTKMMGRLEAERLLVPHGWPAFAPTTPGD  
DA  
GGGTAAPQTCGIVKRLKLAATEQQGTTTPAIAALMQDASVQTPLPVYRITMSPTGQAFAAAARDWA

35

RV  
TRDARPEATVVADAAAPEPGALGRRLTRRICARGPAPPPGRPGRRGPDVREPQDRLQRRAGRYEHH  
PG  
SGHRPEGARPLSPAPRGPGSL\*

40

Gene matched: gi|136939|sp|P10236|UL52\_HSV11  
Gene name: DNA HELICASE/PRIMASE COMPLEX

45

[SEQ ID NO:286] = 15  
ORF # = 57f

ORF start site = 100558  
 ORF end site = 102942  
 ORF sequence:

5 VM  
 GAGKSALTARASCSRG SXSEGGAAARIISYCCSSGRVPQPHSTPSRDAIPEHARGSAPAFPHPTPSG  
 FA  
 GAMGTEDCDHEGRSVAAPVEVMALYATDGCVITSSLALLTNCLLGAEPLYIFS YDAYRPDAPNGPTGA  
 PT

10 EQERFEGSRALYRDAGGLNGDSFRVTFCLLGTEVGVTHHPKGRTRPMFVCRFERADDDAVLQDALGRG  
 TP  
 LLPAHITATLDLEATFALHANIIMALTVAIVHNAPARIGSGSTAPLYEPGESMRSVVGRMSLGQRGLT  
 TL  
 FVHHEARVLAAYRRAYYGSQAQSPFWFLSKFGPDEKSLVLAARYVVLQAPRLGGAGATYDLQAVKDICA

15 TY  
 AIPHDPRPDTLSAASLTSAAITRFCCTSQYSRGAAAAGFPLYVERRIAADVRETGALEKFAHDRSC  
 LR  
 VSDREFITYIYLAHFECFSPRLATHLRAVTTHTDPSAASTEQPSPLGREAVEQFFRHVRAQLNIREY  
 VK

20 QNVTPRETALAGDAAAAYLRARTYAPAALTPAPAYCGVADSSTKMMGRLEAEERLLVPHGWPAPFAPTT  
 PG  
 DDAGGGTAAPQTCGIVKRLKLAATEQQGTPPAIAALMQDASVQTPLPVYRITMSPTGQAFAAAARD  
 DW  
 ARVTRDARPPEATVVADAAAPEPGALGRRLTRRICARGPAPPPGRRGPDVREPQORDLQRRAGRY

25 EH  
 HPGSGHRPEGARPLSPAPRGPGSL\*

Gene matched: gi|136939|sp|P10236|UL52\_HSV11  
 Gene name: DNA HELICASE/PRIMASE COMPLEX

30

[SEQ ID NO:287] = 15  
 ORF # = 57g  
 ORF start site = 100543  
 35 ORF end site = 102942  
 ORF sequence:

MYICRMVAMHVSARRRILSRCAATAPSMVEPSSPGWWRASLSRLTMQAWYVRARARAFTRRRVSSSDS  
 RA  
 SSSVMGAGKSALTARASCSRG SXSEGGAAARIISYCCSSGRVPQPHSTPSRDAIPEHARGSAPAFPH

40 PT  
 PSGFAGAMGTEDCDHEGRSVAAPVEVMALYATDGCVITSSLALLTNCLLGAEPLYIFS YDAYRPDAPN  
 GP  
 TGAPTEQERFEGSRALYRDAGGLNGDSFRVTFCLLGTEVGVTHHPKGRTRPMFVCRFERADDDAVLQD  
 AL

GRGTPLLPAHITATLDLEATFALHANIIMALTVAIVHNAPARIGSGSTAPLYEPGESMRSVVGRMSLG  
 QR  
 GLTTLFVHHEARVLAAYRRAYYGSAQSPFWFLSKFGPDEKSLVLAARYYVLQAPRLGGAGATYDLQAV  
 KD  
 5 ICATYAI PHDPRPDTLSAASLTSFAAITRFCCTSQYSRGAAAAGFPLYVERRIAADVRETGALEKFI  
 HD  
 RSCLRVSDREFITYIYLAHFECFSPRLATHLRAVTTHDPSPAASTEQPSPLGREAVEQFFRHVRAQL  
 NI  
 REYVKQNVTPRETALAGDAAAAYLRARTYAPAALTPAPAYCGVADSSTKMMGRLEAERLLVPHGWPA  
 10 FA  
 PTTPGDDAGGGTAAPQTCGIVKRLCLKLAATEQQGTTTPAIAALMQDASVQTPLPVYRITMSPTGQAF  
 AA  
 ARDDWARVTRDARPEATVVADAAAPEPGALGRRLTRRICARGPAPPPGRRGPDVREPQRDLQR  
 RA  
 15 GRYEHHPGSGHRPEGARPLSPAPRGPGSL\*

Gene matched: gi|136939|sp|P10236|UL52\_HSV11

Gene name: DNA HELICASE/PRIMASE COMPLEX

20 [SEQ ID NO:288] = 15  
 ORF # = 57h  
 ORF start site = 100483  
 ORF end site = 102942  
 25 ORF sequence:  
 MLRMAWETSTSADLSAAPTDMYICRMVAMHVSARRRIILSRCAATAPSMVEPSSPGWWRASLSRLTMQ  
 WY  
 VRARARAFTRRRVSSSDSRASSSVMGAGKSALTARASC SRGSXSEGGAAARIISYCCSSGRVPQPHS  
 TP  
 30 SRDAIPEHARGSAPAFPHPTPSGFAGAMGTEDCDHEGRSVAAPVEVMALYATDGCVITSSLALLTNCL  
 LG  
 AEPLYIFS DAYRPDAPNGPTGAPTEQERFEGSRALYRDAGGLNGDSFRVTFCLLGTEVGVTHHPKGR  
 TR  
 PMFVCRFERADDVAVLQDALGRGTPLLPAHITATLDLEATFALHANIIMALTVAIVHNAPARIGSGST  
 35 AP  
 LYEPGESMRSVVGRMSLGQRGLTTLFVHHEARVLAAYRRAYYGSAQSPFWFLSKFGPDEKSLVLAARY  
 YV  
 LQAPRLGGAGATYDLQAVKDICATYAI PHDPRPDTLSAASLTSFAAITRFCCTSQYSRGAAAAGFPLY  
 VE  
 40 RRIAADVRETGALEKFIHDRSCLRVSDREFITYIYLAHFECFSPRLATHLRAVTTHDPSPAASTEQ  
 PS  
 PLGREAVEQFFRHVRAQLNIREYVKQNVTPRETALAGDAAAAYLRARTYAPAALTPAPAYCGVADSST  
 KM  
 MGRLEAERLLVPHGWPAFAPTTPGDDAGGGTAAPQTCGIVKRLCLKLAATEQQGTTTPAIAALMQDAS  
 45 VQ

TPLPVYRITMSPTGQAFAAAAARDWARVTRDARPEATVVADAAAPEPGALGRRLTRRICARGPAPP  
PG  
RPGRRGPDVREPQDRLQRRAGRYEHHPGSGHRPEGARPLSPAPRGPGSL\*

- 5 Gene matched: gi|136939|sp|P10236|UL52\_HSV11  
Gene name: DNA HELICASE/PRIMASE COMPLEX

- [SEQ ID NO:289] = 15
- 10 ORF # = 57i  
ORF start site = 100242  
ORF end site = 102942  
ORF sequence:  
MTTSLSAMLMAWETSTSDLSAAPTDMYICRMVAMHVSARRRILSRCAATAPSMVEPSSPGWWRASL
- 15 SR  
LTMQAWYVRARARAFTRRRVSSSDSRASSSVMGAGKSALTARASC SRGSXSEGGAAARIISYCCSSG  
RV  
PQPHSTPSRDAIPEHARGSAPAFPHPTPSGFAGAMGTEDCDHEGRSVAAPVEVMALYATDGCVITSSL  
AL
- 20 LTNCLLGAEPLYIFS DAYRPDAPNGPTGAPTEQERFEGSRALYRDAGGLNGDSFRVTFCLLGTEVGV  
TH  
HPKGRTRPMFVCRFERADDVAVLQDALGRGTPLLPAHITATLDLEATFALHANIIMALTVAIVHNAPA  
RI  
GSGSTAPLYEPGESMRSVVGMSLQQRGLTTLFVHHEARVLAAYRRAYYGSASPFWFLSKFGPDEKS
- 25 LV  
LAARYYVLQAPRLGGAGATYDLQAVKDICATYAIPHDPDPDTLSAASLTSAAITRFCCTSQYSRGAA  
AA  
GFPLYVERRIAADVRETGALEKFIANDRSCLRVSDREFITYIYLAHFECFSPRLATHLRAVTTHDPS  
PA
- 30 ASTEQPSPLGREAVEQFFRHVRAQLNIREYVKQNVTPRETALAGDAAAAYLRARTYAPAALTPAPAYC  
GV  
AADSSTKMMGRLAEAERLLVPGWPAFAPPTPGDDAGGGTAAPQTCGIVKRLKLAATEQQGTTTPAIA  
AL  
MQDASVQTPLPVYRITMSPTGQAFAAAAARDWARVTRDARPEATVVADAAAPEPGALGRRLTRRIC
- 35 AR  
GPAPPPGRPGRRGPDVREPQDRLQRRAGRYEHHPGSGHRPEGARPLSPAPRGPGSL\*

Gene matched: gi|136939|sp|P10236|UL52\_HSV11  
Gene name: DNA HELICASE/PRIMASE COMPLEX

40

- [SEQ ID NO:290] = 15  
ORF # = 61a  
ORF start site = 107456  
45 ORF end site = 108016

ORF sequence:

MTTTPLSNLFRLAPDITHVAPPYCLNATWQAENALHTTKTDPACLAARSYLVRASCSTSGPIHCFFFA  
VY  
KDSQHSPLPLVTELRFADLVNHPPVLRELEDKRGGRRLCTGPFSCGTIKDVSGASPAGEYTINGIVYH  
5 CH  
CRYPFSKTCWLGASAAALQHLSISSSGTAARAAEQRRHKIKIKIKV\*

Gene matched: gi|136947|sp|P28281|UL55\_HSV2H

Gene name: PROTEIN UL55^Agi|73806|pir||W

10

[SEQ ID NO:291] = 15

ORF # = 61b

ORF start site = 107372

ORF end site = 108016

15 ORF sequence:

MWGPGPARFIARPGTHGRRVFTDPPPRNMTTTPLSNLFRLAPDITHVAPPYCLNATWQAENALHTTKT  
DP  
ACLAARSYLVRASCSTSGPIHCFFFAVYKDSQHSPLPLVTELRFADLVNHPPVLRELEDKRGGRRLCT  
GP  
20 FSCGTIKDVSGASPAGEYTINGIVYHCHCRYPFSKTCWLGASAAALQHLSISSSGTAARAAEQRRHKI  
KI  
KIKV\*

25 Gene matched: gi|136947|sp|P28281|UL55\_HSV2H

Gene name: PROTEIN UL55^Agi|73806|pir||W

[SEQ ID NO:292] = 15

ORF # = 6a

30 ORF start site = 6446

ORF end site = 8482

ORF sequence:

MAAQRARAPAMRTRGGDAALCAPEDGWVKVHPTPGTMLFREILLGQMGYTEGQGVYNVRSSEATRQ  
LQ  
35 AAIFHALLNATTYRDLEEDWRRHVVARGLQPQRLVRRYRNAREGDIAGVAERVFDTWRCTLRITLLDF  
AH  
GVVDCFAPGGPSGPTSFPKYIDWLTCLGLVPILRKTREGEATQRLGAFLRQHTLPRQLATVAGAAERA  
GP  
GLELAVAFDSTRMAEYDRVHIYNNHRRGEWLVRDPVSGQRGECLVLCPLWTGDRLVFDSFPVQRLCP  
40 EI  
VACHALREHAHICRLNRTASVKVLLGRKSDSERGVAGAAARVVNKALGEDDETKAGSAASRLVRLIIM  
KG  
MRHVGDIINDTVRAYLDEAGGHLIDTPAVDHTLPGFGKGGTGRGSAAQDPGARPPQLRQAFQTAVVNNI  
NG



MLEGYINNLFGTIERLRETNAGLATQLQARDRELRRQAQALEREQRAADRAAGGGAGRPAEADLLRA  
 DY  
 DIIDVSKSMDDDTYVANSFQHQYIPAYGQDLERLSRLWEHELVRCKILRHRNNQGQETSISYSSGAI  
 AS  
 5 FVAPYFEYVLRAPRAGALITGSDVILGEEELWEAVFKKTRLQTYLTDVAALFVADVQHAALPRPPSPT  
 PA  
 DFRASASPRGGSRSRTRTRSRSRSPGRTPRGAPDQGWGVERRDGRPHARR\*

10 Gene matched: gi|136794|sp|P10190|UL06\_HSV11  
 Gene name: VIRION PROTEIN UL6^Agi|73994|

[SEQ ID NO:293] = 15

ORF # = 6b

15 ORF start site = 6326

ORF end site = 8482

ORF sequence:

MDVKFKNASSLNRTAGLAPGCCGGGPGARTSREPSPPDAAMAAQRARAPAMRTRGGDAALCAPEDGWV  
 KV  
 20 HPTPGTMLFREILLGQMGYTEGQGVYNVRSSEAATRLQAAIFHALLNATTYRDLEEDWRRHVVAR  
 LQ  
 PQRLVRRYRNAREGDIAGVAERVFDTWRCTLRITLLDFAHGVVDCFAPGGPSGPTSPFKYIDWLTC  
 LG  
 LV  
 PILRKTREREATQRLGAFLRQHTLPRQLATVAGAAERAGPGLLELAVAFDSTRMAEYDRVHIYNNHRR  
 25 GE  
 WLVRDPVSGQRGECLVLCPLWTGDRLVFDSFVQRLCPEIVACHALREHAHICRLNTASVKVLLGRK  
 SD  
 SERGVAGAARVVKALGEDDETKAGSASRLVRLIINMKGMRHVGDIINDTVRAYLDEAGGHLIDTPAV  
 DH  
 30 TLPFGFGKGGTGRGSAAQDPGARPPQLRQAFQTAVVNNINGMLEGYINNLFGTIERLRETNAGLATQLQ  
 AR  
 DRELRRQAQALEREQRAADRAAGGGAGRPAEADLLRADYDIIDVSKSMDDDTYVANSFQHQYIPAYG  
 QD  
 LERLSRLWEHELVRCKILRHRNNQGQETSISYSSGAIASFVAPYFEYVLRAPRAGALITGSDVILGE  
 35 EE  
 LWEAVFKKTRLQTYLTDVAALFVADVQHAALPRPPSPTPADFRASASPRGGSRSRTRTRSRSRSPGRTPR  
 GA  
 PDQGWGVERRDGRPHARR\*

40 Gene matched: gi|136794|sp|P10190|UL06\_HSV11  
 Gene name: VIRION PROTEIN UL6^Agi|73994|

[SEQ ID NO:294] = 15

ORF # = 6c

45 ORF start site = 6296

ORF end site = 8482  
ORF sequence:  
MRAMIGWTPCMDVKFKNASSLNRTAGLAPGCCGGGPGARTSREPSPPDAAMAAQRARAPAMRTRGGDA  
AL  
5 CAPEDGWVKVHPTPGTMLFREILLGQMGYTEGQGVYNVVRSSSEATRQLQAAIFHALLNATTYRDLEE  
DW  
RRHVVARGLQPQRLVRRYRNAREGDIAGVAERVFDTWRCTLRITLLDFAHGVVDCFAPGGPSGPTSFP  
KY  
IDWLTCLGLVPILRKTREREATQRLGAFRLRQHTLPRLQATVAGAAERAGPGLLELAVAFDSTRMAEYD  
10 RV  
HIYYNHRRGEWLVRDPVSGQGECLVLCPLWTGDRLVFDSVPVQRLCPEIVACHALREHAHICRLRNT  
AS  
VKVLLGRKSDSERGVAGAAARVVNKALGEDDETAKGSAASRLVRLIINMKGMRHVGDINDTVRAYLDEA  
GG  
15 HLIDTPAVDHTLPGFGKGGTGRGSAQDPGARPPQLRQAFQTAVVNNINGMLEGYINNLFGTIERLRE  
TN  
AGLATQLQARDRELRRQAQALEREQRAADRAAGGGAGRPAEADLLRADYDIIDVSKSMDDDTYVANS  
FQ  
HQYIPAYGQDLERLSRLWEHELVRCKILRHRNNQGETSISYSSGAIASFVAPYFEYVLRAPRAGAL  
20 IT  
GSDVILGEEELWEAVFKKTRLQTYLTDVAALFVADVQHAALPRPPSPTPADFRASASPRGGSRSRTRT  
RS  
RSPGRTPRGAPDQGWGVERRDGRPHARR\*

25 Gene matched: gi|136794|sp|P10190|UL06\_HSV11  
Gene name: VIRION PROTEIN UL6^Agi|73994|

[SEQ ID NO:295] = 15  
ORF # = 6d  
30 ORF start site = 6167  
ORF end site = 8482  
ORF sequence:  
MRYAANGNSRSGRPVGTSKAATSRNHCRRGTCVTSSCCCESSMRAMIGWTPCMDVKFKNASSLNRTA  
GL  
35 APGCCGGGPGARTSREPSPPDAAMAAQRARAPAMRTRGGDAALCAPEDGWVKVHPTPGTMLFREILLG  
QM  
GYTEGQGVYNVVRSSSEATRQLQAAIFHALLNATTYRDLEEDWRRHVVARGLQPQRLVRRYRNAREGD  
IA  
GVAERVFDTWRCTLRITLLDFAHGVVDCFAPGGPSGPTSFPKYIDWLTCLGLVPILRKTREREATQRL  
40 GA  
FLRQHTLPRLQATVAGAAERAGPGLLELAVAFDSTRMAEYDRVHIYYNHRRGEWLVRDPVSGQGECL  
VL  
CPPLWTGDRLVFDSVPVQRLCPEIVACHALREHAHICRLRNTASVKVLLGRKSDSERGVAGAAARVVNKA  
LG

EDDETKAGSAASRLVRLIINMKGMRHVGDINDTVRAYLDEAGGHLIDTPAVDHTLPGFGKGGTGRGSA  
 AQ  
 DPGARPQQLRQAFQTAVVNNINGMLEGYINNLFGTIERLRETNAGLATQLQARDRELRAQAGALERE  
 QR  
 5 AADRAAGGAGRPAEADLLRADYDIIDVSKSMDDDTYVANSFQHQYIPAYGQDLERLSRLWEHELVR  
 FK  
 ILRHRNNQGQETSISYSSGAIASFVAPYFEYVLRAPRAGALITGSDVILGEEELWEAVFKKTRLQTYL  
 TD  
 VAALFVADVQHAALPRPPSPTPADFRASASPRGGSRSRTRTRSRSRPGRTPRGAPDQGWGVERRDGRPH  
 10 AR  
 R\*

Gene matched: gi|136794|sp|P10190|UL06\_HSV11

Gene name: VIRION PROTEIN UL6^Agi|73994|

15 [SEQ ID NO:296] = 15  
 ORF # = 6e  
 ORF start site = 6065  
 ORF end site = 8482  
 20 ORF sequence:  
 MFCAAIRVAPVTTQSRSTSLRVCTHVLPDPALPVMRYAANGNSRSGRPVGTSKAATSRNHCRRGTCVT  
 SS  
 CCESSSRMRAMIGWTPCMDVKFNASSLNRTAGLAPGCCGGGPGARTSREPSPPDAAMAAQRARAPAM  
 RT  
 25 RGGDAALCAPEDGWVKVHPTPGTMLFREILLGQMGYTEGQGVYNVRSSEAATRQLQAAIFHALLNAT  
 TY  
 RDLEEDWRRHVVARGLQPQRLVRRYRNAREGDIAGVAERVFDTWRCTLRITLLDFAHGVVDCFAPGGP  
 SG  
 PTSFPHYIDWLTCLGLVPILRKTREREATQRLGAFLRQHTLPRQLATVAGAAERAGPGLLELAVAFDS  
 30 TR  
 MAEYDRVHIYYNHRERGEWLVRDPVSGQRGECLVLCPLWTGDRLVFDSPVQRLCPEIVACHALREHAH  
 IC  
 RLRNTASVKVLLGRKSDSERGVAGAARVVNKALGEDDETAKAGSAASRLVRLIINMKGMRHVGDINDTV  
 RA  
 35 YLDEAGGHLIDTPAVDHTLPGFGKGGTGRGSAAQDPGARPQQLRQAFQTAVVNNINGMLEGYINNLF  
 TI  
 ERLRETNAGLATQLQARDRELRAQAGALEREQRAADRAAGGAGRPAEADLLRADYDIIDVSKSMDD  
 DT  
 YVANSFQHQYIPAYGQDLERLSRLWEHELVRCKILRHRNNQGQETSISYSSGAIASFVAPYFEYVLR  
 40 AP  
 RAGALITGSDVILGEEELWEAVFKKTRLQTYLTDVAALFVADVQHAALPRPPSPTPADFRASASPRGG  
 SR  
 SRTTRTRSRSRPGRTPRGAPDQGWGVERRDGRPHARR\*

45 Gene matched: gi|136794|sp|P10190|UL06\_HSV11

Gene name: VIRION PROTEIN UL6^Agi|73994|

[SEQ ID NO:297] = 15  
 5 ORF # = 6f  
 ORF start site = 6026  
 ORF end site = 8482  
 ORF sequence:  
 MGR LRNAPESLT YMFC AAIRVAPVTTQSRTSLRVCTHVLFPDPALPVMRYAANGNSRSGRPVGTSKAA  
 10 TS  
 RNHCRRGTCVTSSCCCESSRM RAMIGWTPCMDVKFKNASSLNRTAGLAPGCCGGGPGARTSREPSPPD  
 AA  
 MAAQRARAPAMRTRGDAALCAPEDGWVKVHPTPGTMLFREILLGQMGYTEGQGVYNVVRSSAATRQ  
 LQ  
 15 AAIFHALLNATTYRDLEEDWRRHVVARGLQPQRLVRRYRNAREGDIAGVAERVFDTWRCTLR TLLDF  
 AH  
 GVVD CFAPGGPSGPTSF PKYIDWLTCLGLVPILRK TREGEATQRLGAFLRQH TLPRLATVAGAAERA  
 GP  
 GLELAVAFDSTRMAEYDRVHIYYNHRRGEWLVRDPVSGQRGECLVLC PPLWTGDRLVFDSPVQRLCP  
 20 EI  
 VACHALREHAHICRLNTASVKVLLGRKSDSERGVAGAARVVKALGEDDETKAGSAASRLVRLIINM  
 KG  
 MRHVGDINDTVRAYLDEAGGHLIDTPAVDHTLPGFGKGGTGRGSAAQDPGAR PQQLRQAFQTAVVNNI  
 NG  
 25 MLEGYINNLFGTIERLRETNAGLATQLQARDRELRR AQAGALEREQRAADRAAGGGAGRP AEADLLRA  
 DY  
 DIIDVSKSMDDDTYVANSFQH QYIPAYGQDLERLSRLWEHELVR CFKILRHRNNQGQETSISYSSGAI  
 AS  
 FVAPYFEYVLRAPRAGALITGSDVILGEEELWEAVFKKTRLQTYLTDVAALFVADVQHAALPRPPSPT  
 30 PA  
 DFRASASPRGGSRSRTRTRSRSPGRTPRGAPDQGWGVERRDGRPHARR\*

Gene matched: gi|136794|sp|P10190|UL06\_HSV11

Gene name: VIRION PROTEIN UL6^Agi|73994|

35 [SEQ ID NO:298] = 15  
 ORF # = 6g  
 ORF start site = 6017  
 ORF end site = 8482  
 40 ORF sequence:  
 MVL MGR LRNAPESLT YMFC AAIRVAPVTTQSRTSLRVCTHVLFPDPALPVMRYAANGNSRSGRPVGTS  
 KA  
 ATSRNHCRGTCVTSSCCCESSRM RAMIGWTPCMDVKFKNASSLNRTAGLAPGCCGGGPGARTSREPS  
 PP

DAAMAAQRARAPAMRTRGGDAALCAPEDGWVKVHPTPGTMLFREILLGQMGYTEGQGVYVVRSEAA  
 TR  
 QLQAAIFHALLNATTYRDLEEDWRRHVVARGLQPQRLVRRYRNAREGDIAGVAERVFDTWRCTLRRTL  
 LD  
 5 FAHGVVDCFAPGGPSGPTSFPKYIDWLTCLGLVPILRKTREREATQRLGAFRLQHTLPRQLATVAGAA  
 ER  
 AGPGLLELAVAFDSTRMAEYDRVHIYNNHRRGEWLVRDPVSGQRGECLVLCPLWTGDRLVFDSVPVQR  
 LC  
 PEIVACHALREHAHICRLRNTASVKVLLGRKSDSERGVAGAARVVNKALGEDDETKAGSAASRLVRLI  
 10 IN  
 MKGMRHVGDIINDTVRAYLDEAGGHLIDTPAVDHTLPGFGKGGTGRGSAAQDPGARPQQLRQAFQTAVV  
 NN  
 INGMLEGYINNLFGTIERLRETNAGLATQLQARDRELRAQAGALEREQRAADRAAGGGAGRPAEADL  
 LR  
 15 ADYDIIIDVSKSMDDDTYVANSFQHQYIPAYGQDLERLSRLWEHELVRCKILRHRNNQGQETSISYSS  
 GA  
 IASFVAPYFEYVLRAPRAGALITGSDVILGEEELWEAVFKKTRLQTYLTDVAALFVADVQHAALPRPP  
 SP  
 TPADFRASASPRGGSRSRTRTRSRSRPGRTPRGAPDQGWGVERRDGRPHARR\*

20

Gene matched: gi|136794|sp|P10190|UL06\_HSV11

Gene name: VIRION PROTEIN UL6^Agi|73994|

25 [SEQ ID NO:299] = 15

ORF # = 47a

ORF start site = 88122

ORF end site = 89564

ORF sequence:

30 MALGRVGLAVGLWGLLVGVVVVLANASPGRTITVGPRGNASNAAPSASPRNASAPRTTPTPPQPRKA  
 TK  
 SKASTAKPAPPPKTGPPKTSSEPVRCNRHDPLARYGSRVQIRCFPNSTRTESRLQIWRYATATDAEI  
 GT  
 APSLEEVMVNVSAPPGQLVYDSAPNRTDPHVIWAEGAGPGASPRLYSVVGPLGRQRLIIIEELTLETQ  
 35 GM  
 YYVWVGRTDRPSAYGTWVRVRVFRPPSLTIHPHAVLEGQPFKATCTAATYYPGNRAEFVWFEDGRRVF  
 DP  
 AQIHTQTQENPDGFSTVSTVTSAAVGGQGPRTFTCQLTWHRDSVSFSRRNASGTASVLPRTITMEF  
 TG  
 40 DHAVCTAGCVPEGVTFWFLGDDSSPAEKVAVASQTSCGRPGTATIRSTLPVSYEQTEYICRLAGYPD  
 GI  
 PVLEHHGSHQPPRPDPTERQVIRAVEGAGIGAVLVAVVLAGTAVVYLTHASSVRYRRLR\*

Gene matched: gi|138220|sp|P06475|VGLC\_HSV23

45 Gene name: GLYCOPROTEIN C PRECURSOR^Agi|

[SEQ ID NO:300] = 15  
ORF # = 47b  
5 ORF start site = 87918  
ORF end site = 89564  
ORF sequence:  
MGAGVPWTGIKARGAGGPITVRVLGWEVAQKATHPCCSCPREAVVSGNPPRCAGRAHRSFAGAGALLV  
MA  
10 LGRVGLAVGLWGLLWVGVVVLANASPGRTITVGPRGNASNAAPSASPRNASAPRTTTPPQPRKATK  
SK  
ASTAKPAPPPKTGPPKTSSEPVRNCRHDPLARYGSRVQIRCFPNSTRTESRLQIWRATATDAEIGT  
AP  
SLEEVNVNSAPPGQLVYDSAPNRTPHVIWAEGAGPGASPRLYSVVGPLGRQRLIEELTLETQGM  
15 YY  
VWGRTRDRPSAYGTWVRVRVFRPPSLTIHPHAVLEGQPFKATCTAATYYPGNRAEFVWFEDGRRVFD  
AQ  
IHTQTQENPDGFSTVSTVTSAAVGGQGPRTFTCQLTWHRDSVSFSRRNASGTASVLPRPTITMEFTG  
DH  
20 AVCTAGCVPEGVTFAWFLGDDSSPAEKVAVASQTSCGRPGTATIRSTLPVSYEQTEYICRLAGYPDGI  
PV  
LEHHGSHQPPPRDPTERQVIRAVEGAGIGVAVLVAVVLGTA VVYLTHASSVRYRRLR\*

25 Gene matched: gi|138220|sp|P06475|VGLC\_HSV23  
Gene name: GLYCOPROTEIN C PRECURSOR^Agi|

[SEQ ID NO:301] = 15  
30 ORF # = 52a  
ORF start site = 97076  
ORF end site = 95441  
ORF sequence:  
MSVLGDARHPRRFPSPRGPRPFSVAGPGSLPPSPPPGARARLIRLSRSLFPDPTAPMDLLVDDL FADAD  
35 GV  
SPPPPRPAGGPKNTPAAPPLYATGRLSQAQLMPSPMPVPPAALFNRLDDLGFSAGPALCTMLDTWN  
ED  
LFSGFPTNADMYRECKFLSTLPDVIDWGDHVPERSPIDIRAHGDVAFPTLPATRDELPSYYEAMAQ  
FF  
40 RGE LRAREESYRTVLANFCSALYRYLRASVRQLHRQAHRGRNRDLREMLRTTIADRYRETARLARV  
LF  
LHLYLFLSREILWAAYAEQMMRPDLFDGLCCDLESWRQLACLFQPLMFINGSLTVRGVPVEARRLREL  
NH  
IREHLNLPLVRSAAAEPGA PLTTPV LQGNQARSSGYFMLLIRAKLDSYSSVATSEGESVMREHAYS  
45 RG

RTRNNYGSTIEGLLDLPDDDDAPAEAGLVAPRMSFSLAGQRPRLSTTAPITDVS LGDELRLDGEEVD  
 MT  
 PADALDDFDLEMLGDVESPSPGMTHDPVSYGALDVDDFEFEQMFTDAMGIDDFGG\*

- 5 Gene matched: gi|1168549|sp|P29793|ATIN\_HSV23  
 Gene name: ALPHA TRANS-INDUCING PROTEIN

- [SEQ ID NO:302] = 15  
 10 ORF # = 52b  
 ORF start site = 969103  
 ORF end site = 95441  
 ORF sequence:  
 MDLLVDDLFADADGVSPPPRPAGGPKNTPAAPPLYATGRLSQAQLMPSPMPVPPAALFNRLDDL  
 15 FS  
 AGPALCTMLDTWNEDLFSGFPTNADMYRECKFLSTLPSDVIDWGAHVPERSPIDIRAHGDVAFPTLP  
 AT  
 RDELPSYYEAMAQFFRGELRAREESYRTVLNFCALYRYLRASVRQLHRQAHRGRNRDLREMLRTT  
 IA  
 20 DRYRETARLARVFLHLYLFLSREILWAAAEQMMRPDLFDGLCCDLESWRQLACLFQPLMFINGSL  
 TV  
 RGVPEARRLRELNHIREHLNPLVRSAAAEPPGAPLTTPPVLQGNQARSSGYFMLLIRAKLDSYSSV  
 AT  
 SEGESVMREHAYSRRGRTRNNYGSTIEGLLDLPDDDDAPAEAGLVAPRMSFSLAGQRPRLSTTAPITD  
 25 VS  
 LGDELRLDGEEVDMTPADALDDFDLEMLGDVESPSPGMTHDPVSYGALDVDDFEFEQMFTDAMGIDDF  
 GG\*

- Gene matched: gi|1168549|sp|P29793|ATIN\_HSV23  
 30 Gene name: ALPHA TRANS-INDUCING PROTEIN

- [SEQ ID NO:303] = 15  
 ORF # = 52c  
 35 ORF start site = 97097  
 ORF end site = 95441  
 ORF sequence:  
 MRGGGREMSVLGDARHPRRFPSPRGPRPFSVAGPGSLPPSPPPGARARLIRLSRSLFPDPTAPMDLLVD  
 DL  
 40 FADADGVSPPPRPAGGPKNTPAAPPLYATGRLSQAQLMPSPMPVPPAALFNRLDDLGFSA  
 GPALC  
 TM  
 LDTWNEDLFSGFPTNADMYRECKFLSTLPSDVIDWGAHVPERSPIDIRAHGDVAFPTLPATRDELPS  
 YY  
 EAMAQFFRGELRAREESYRTVLNFCALYRYLRASVRQLHRQAHRGRNRDLREMLRTTIADRYRE  
 45 TA

RLARVLFHLHYLFLSREILWAAYAEQMMRPDLFDGLCCDLESWRQLACLFQPLMFINGS�TVRGVPVE  
AR  
RLRELNHIREHLNPLVRSAAAEPEGAPLTPPVQLQGNQARSSGYFMLLIRAKLSYSSVATSEGESV  
MR  
5 EHAYSRGRTRNNYGSTIEGLLDLPDDDDAPAEAGLVAPRMSFSLSAGQRPRLSTTAPITDVSLGDELR  
LD  
GEEVDMTPADALDDFDLEMLGDVESPSPGMTHDPVSYGALDVDDFEFEQMFTDAMGIDDFGG\*

- 10 Gene matched: gi|1168549|sp|P29793|ATIN\_HSV23  
Gene name: ALPHA TRANS-INDUCING PROTEIN



## SEQUENCE LISTING

## (1) GENERAL INFORMATION

- 5 (i) APPLICANT: SmithKline Beecham Corporation
- (ii) TITLE OF THE INVENTION: Novel Coding Sequences from Herpes Simplex Virus Type-2
- 10 (iii) NUMBER OF SEQUENCES: 303
- (iv) CORRESPONDENCE ADDRESS:
- (A) ADDRESSEE: SmithKline Beecham Corporation
- (B) STREET: 709 Swedeland Road
- 15 (C) CITY: King of Prussia
- (D) STATE: PA
- (E) COUNTRY: U.S.A.
- (F) ZIP: 19046
- 20 (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Diskette
- (B) COMPUTER: IBM Compatible
- (C) OPERATING SYSTEM: Windows
- (D) SOFTWARE: FastSEQ for Windows Version 2.0b
- 25 (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:
- 30 (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER: 60/049,018
- (B) FILING DATE: 09-JUN-1997
- 35 (A) APPLICATION NUMBER: 60/030,279
- (B) FILING DATE: 04-NOV-1996
- (viii) ATTORNEY/AGENT INFORMATION:
- 40 (A) NAME: Geiger, Kathleen W.
- (B) REGISTRATION NUMBER: 35,880
- (C) REFERENCE/DOCKET NUMBER: P50583

## (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 610-270-5968

(B) TELEFAX: 610-270-5090

(C) TELEX:

5

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 8953 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

	GCGGTCGATC TAGAGGATCC CCCGCGCGCC TGCAGCTGGG TCGCCAGACC CGCGTTCGTC	60
	TCTCGCAGGC GTTCTATGGT TCCAAAGAGA TTATTGATAT AGCCCTCCAG CATGCCGTTG	120
20	ATGTTGTGTA CCACGCCCGT CTGAAACGCC TGGCGAAGCT GCTGCGGTCG CGCCCCCGGG	180
	TCCTGGGCCG CCGACCCGCG GCCGGTGCCG CCCTTGCCGA ACCCAGGGAG GGTGTGGTCG	240
	ACGGCGGGGG TGTCGATCAG GTGCCCCCCC GCCTCGTCCA AGTAGGCGCG TACCGTGTCTG	300
	TTGATGTCTG CCACGTGGCG CATGCCCTTC ATGTTGATGA TGAGCCGCAC GAGACACGAG	360
	GCGGCCGAGC CGGCCTTCGT CTCGTCTATC TCCCCAGCG CCTTATTGAC GACCCGCGCG	420
25	GCGCCAGCCA CCCC GCGCTC GCTGTCTGCTC TTGCGCCCCA ACAGCACCTT GACGGACGCG	480
	GTGTTGCGCA GACGGCAGAT GTGCGCGTGT TCCCGGAGGG CGTGGCACGC GACGATCTCG	540
	GGGCACAGCC GCTGAACGGG CGAATCGAAG ACCAGGCGGT CGCCGGTCCA CAGGGGGGGG	600
	CACAGACCA GGCACCTCGC GCGCTGCCCG CTGACCGGGT CGCGCACCAG CCACTCCCCC	660
	CGGCGATGGT TGTAAGTAGAT GTGCACACGG TCGTATTCCG CCATGCGCGT GGAGTCGAAC	720
30	GCGACGCCA GCTCCAGAAG CCCC GGGCCG GCGCGTCCG CGGCCCCGGC GACCGTGGCC	780
	AGCTGCCGGG GCAGCGTGTG CTGCCTGAGA AACGCCCCCA GGCCTGCGT CGCCTCCCCC	840
	TCGCGCGTCT TGC GCAATAT GGAACCAGC CCCAGACACG TCAGCCAGTC GATATATTTG	900
	GGGAAGCTGG TCGGTCCGCT TGGGCCGCCC GGC GCAAAGC AGTTTACCAC CCCGTGGGCA	960
	AAGTCCAGCA GCGTCGTCTT GAGCGTGCAT CGCCACGTGT CGAACACCCG CTCGGCCACC	1020
35	CCGGCGATAT CGCCCTCCCG GCGGTTCCGG TACCTGCGAA CCAGCCGCTG CGGCTGGAGG	1080
	CCGCGGGCCA CCACGTGGCG GCGCCAGTCC TCCTCCAGGT CCCGCTACGT TGTGGCGTTG	1140
	AGGAGCGCGT GGAAGATCGC CGCCTGCAGC TGTCGGGTGG CGGCCTCGCT GGACCGGACG	1200
	ACGTTGTACA CCCCCTGACC CTCGGTGTAC CCCATCTGCC CGAGGAGAAT CTCGCGGAAC	1260
	AACATCGTCC CGGGGGTGGG GTGAACCTTC ACCCAGCCGT CCTCGGGGGC GCATAGCGCC	1320
40	GCGTCGCCG CCCGCGTCCG CATCGCCGGC GCGCGCGCG GCTGTGCGGC CATGGCGGCG	1380
	TCCGGCGGGG AGGGCTCGCG GGACGTCCGG GCACCAAGTC CGCCCCACA GCAGCCCGGG	1440
	GCCAGACCCG CCGTGCGGTT CAGGGACGAG GCGTTTTTAA ATTTTACGTC CATGCACGGG	1500
	GTCCAACCAA TCATCGCACG CATCCGAGAG CTCTCGCAGC AGCAGCTCGA CGTCACGCAG	1560

	GTGCCGCGCC	TGCAGTGGTT	CCGGGACGTG	GCGGCCTTGG	AGGTCCCGAC	CGGCCTGCCG	1620
	CTCCGGGAGT	TTCCGTTCGC	GGCGTATCTC	ATCACCGGCA	ACGCCGGATC	CGGAAAGAGT	1680
	ACGTGCGTGC	AGACCCTCAA	CGAGGTCCCTG	GACTGCGTGG	TCACGGGCGC	CACGCGAATC	1740
	GCGGCGCAGA	ACATGTACGT	TAAGCTCTCG	GGGGCGTTTC	TGAGTCGACC	CATCAACACC	1800
5	ATCTTTCACG	AGTTCGGGTT	TCGCGGGAAT	CACGTCCAGG	CCCAGCTGGG	GCAGCACCCG	1860
	TACACCCTGG	CCAGCAGCCC	CGCCTCGCTG	GAAGACCTGC	AGCGGCGAGA	CCTGACGTAC	1920
	TACTGGGAGG	TGATCCTCGA	CATCACCAAG	CGGGCCCTGG	CGGCGCACGG	GGGCGAAGAC	1980
	GCGCGAAACG	AGTTCACGCG	CCTCACCGCC	CTAGAGCAGA	CTTTGGGGCT	GGGCCAGGGT	2040
	GCCCTCACGC	GCCTGGCCTC	GGTCACACAC	GGGGCGCTGC	CGGCTTTTAC	CCGCAGCAAC	2100
10	ATTATCGTCA	TCGACGAGGC	CGGGCTCCTG	GGGCGGCACC	TACTCACGAC	CGTGGTGTAT	2160
	TGCTGGTGGA	TGATTAACGC	CCTGTACCAC	ACCCCCAGT	ACGCGGGCCG	CCTGCGGCCG	2220
	GTGTGGTGT	GCGTGGGGTC	GCCGACCCAG	ACGGCCTCGC	TGGAGTCCAC	CTTCGAACAC	2280
	CAAAAACCTGC	GATGCTCCGT	CCGGCAGAGC	GAAAACGTGC	TCACGTACCT	CATCTGCAAC	2340
	CGCACCCCTAC	GCGAGTACAC	GCGCCTCTCG	CACAGCTGGG	CCATTTTCAT	TAACAACAAG	2400
15	CGATGTGTGG	AGCACGAGTT	CGGGAACCTC	ATGAAGGTGC	TGGAGTACGG	CCTTCCCATC	2460
	ACCGAGGAGC	ACATGCAGTT	TGTGGACCGC	TTTGTCTGCC	CGGAAAGTTA	CATCACCAAC	2520
	CCGGCCAACC	TTCCGGGGTG	GACGCGGCTG	TTCTCGTCCC	ACAAGGAGGT	CAGCGCGTAC	2580
	ATGGCCAAGC	TCCACGCCTA	CCTAAAGGTG	ACTCGCGAGG	GGGAGTTTGT	TGTGTTTACC	2640
	CTCCCCGTGC	TTACGTTTGT	GTCGGTCAAA	GAGTTTGACA	AGTATCGACG	GCTCACGCAG	2700
20	CAACCCACGC	TGACCATGGA	AAAGTGGATC	ACGGCCAACG	CCAGTCGCAT	CACCAACTAC	2760
	TCCCAGAGTC	AGGACCAGGA	CGCGGGGCAC	GTGCGCTGTG	AGGTGCACAG	CAAGCAACAG	2820
	CTAGTCGTGG	CCCGGAACGA	CATCACGTAC	GTCCTCAACA	GCCAGGTTCG	GGTGACCGCG	2880
	CGCTCCGAA	AGATGGTGT	TGGGTTCGAC	GGGACGTTTC	GGACCTTCGA	GGCTGTGCTG	2940
	CGCGACGACA	GCTTCGTGAA	GACCCAGGGG	GAGACCTCGG	TGGAGTTCGC	CTACCGGTTC	3000
25	CTGTCGCGGC	TCATGTTTCGG	CGGGCTGATT	CACTTTTACA	ACTTCTCCA	GCGCCCCGGC	3060
	CTGGACGCGA	CCCAGAGGAC	CCTGGCCTAC	GGCCGCCTAG	GGGAGCTGAC	GGCAGAACTC	3120
	CTGTCGCTAC	GCCGGGACGC	CGCCGGCGCA	TCGGCAACCA	GGGCCGCCGA	CACCAGCGAC	3180
	CGCTCTCCGG	GGGAGCGTGC	GTTCAATTTT	AAGCACCTGG	GCCCGCGGGA	CGGGGGCCCG	3240
	GACGACTTCC	CCGACGACGA	CCTTGACGTT	ATCTTCGCCG	GGCTGGACGA	ACAGCAGCTG	3300
30	GACGTGTTCT	ACTGCCACTA	CGCCCTCGAA	GAGCCGGAGA	CCACCGCGGC	CGTCCACGCC	3360
	CAGTTTGGGC	TCCTGAAGAG	GGCCTTTCTG	GGGCGATACC	TTATCCTACG	GGAGCTCTTC	3420
	GGGAGGTGT	TTGAGAGCGC	CCCCTTCAGC	ACCTACGTGG	ACAATGTCAT	CTTCCGGGGC	3480
	TGCGAGCTGC	TGACCGGCTC	GCCGCGCGGG	GGGCTGATGT	CCGTGGCCCT	GCAGACGGAC	3540
	AACTACACGC	TGATGGGGTA	CACGTACACC	CGGGTGTTCG	CGTTTCGCGA	GGAGCTGCGG	3600
35	CGGCGGCACG	CGACGGCCGG	CGTGGCCGAG	TTCTTGAGAG	AGTCCCCCCT	GCCCTACATC	3660
	GTCCTGCGGG	ACCAGCACGG	CTTCATGTCT	GTCGTCAATA	CCAACATCAG	TGAGTTTGTC	3720
	GAGTCGATCG	ACTCCACGGA	GCTGGCCATG	GCCATCAACG	CCGACTACGG	CATCAGCTCC	3780
	AAACTCGCGA	TGACCATCAC	GCGCTCCCAG	GGGCTCAGTC	TGGACAAGGT	CGCCATATGC	3840
	TTCACGCCCG	GAAACCTGCG	CCTAAACAGC	GCGTACGTAG	CCATGTCCCG	CACCACCTCA	3900
40	TCCGAGTTCC	TGCACATGAA	TCTAAACCCG	CTCCGGGAGC	GCCACGAACG	CGATGACGTC	3960
	ATTAGCGAGC	ACATACTATC	TGCTCTACGC	GATCCGAATG	TGGTCATTGT	CTATTAACCC	4020
	TCCATTCCCT	CGCGTCCCA	CCGCACCCGG	GCCGGGTGAC	ATTACCCCCC	ACCCCCCGA	4080
	GACATGGGGA	ACCCCCAGAC	GACCATCGCG	TACAGCCTAC	ATCACCCAG	GGCGTCGCTA	4140

	ACGAGCGCGC	TGCCGGACGC	GGCACAGGTG	GTGCACGTGT	TTGAGTCAGG	GACGCGCGCG	4200
	GTTCTGACGC	GGGGTCGAGC	GCGCCAGGAC	CGCCTGCCCC	GCGGAGGCGT	GGTGATACAA	4260
	CACACCCCA	TCGGGCTGCT	GGTGATTATC	GACTGTCTGT	CCGAATTTTG	CGCATACCGC	4320
	TTTATAGGAC	GCGCTAGTAC	CCAGAGGCTG	GAGCGCTGGT	GGGACGCCCA	TATGTACGCG	4380
5	TACCCCTTTG	ACTCCTGGGT	CAGCTCATCG	CACGGCGAAA	GCGTCCGGAG	TGCGACGGCC	4440
	GGCATCCTGA	CGGTGGTGTG	GACCCCGGAC	ACCATCTACA	TCACCGCAAC	GATCTACGGG	4500
	ACGGCCCCCG	AGGCGGCGCG	GGGGTGCAT	AACGCACCCC	TGGACGTCCG	CCCAACCACA	4560
	CCCCCGCCCC	CCGTATCCCC	AACGGCGGGC	GAGTTCCCAG	CAAACACAAC	AGACCTACTG	4620
	GTGAGGTTT	TGCGGGAAAT	TCAGATCAGC	CCCACCTTGG	ACGACGCAGA	CCCAACCCCC	4680
10	GGAACCTGAA	ACCTTCTTTC	CTCCCCACCC	CGCCCGCTTG	CATATTCCCT	CTGCGCGCGG	4740
	CGACGGCACC	GCCGGGCGAA	CGAACGGTCA	ATAAAATCAA	TCAATCCATC	ATCCAACAAA	4800
	ATAAGCTACG	TGTTATTTAT	TGAAACGTCA	CAACACATCA	GTAACGGGGG	GAAGGGTAGG	4860
	GGGAAAAGAA	AGGGGACGGC	GGGGGTGCTT	AGTCTGGTTC	CGTAGACAGC	ATGACGTTAT	4920
	CTCGATGGAG	GCGCATGGGT	TCGGACGAAA	CAAACCTCGT	TACAAACACG	GGGACGGCCG	4980
15	GGCAGAGCCG	CCCATCCGAG	GACCGCGTGT	AATACTTGCG	CGGCTTGCGC	GACCGAATAA	5040
	CCGCCCCGAG	CTGCTCGCGC	ACCTGCGCGG	CGTTGGCGCG	CTTGCACAAA	ACGAACATCT	5100
	GGAGGCTCTT	GTTGCTGCGT	GGTACGCATG	TGCGTTGCGG	CGGTGCTCCG	CGCTTGCGCT	5160
	GGCGCGCGGC	CGTCCCCGAA	AACGACGAGG	TCTTGGTACA	CGCGATGCTG	AACTTGGCCA	5220
	GCGACAGCCG	CAGGTCCTTA	CGGATCGTAT	CCGTGAGCTG	GCGGCGCCCC	AGTTTCGTCGA	5280
20	TCGAAGATAC	CATAAACAGA	GTATCAAAGG	TGACGTAGGG	GCCGTCCGCC	CCCGGCGGGT	5340
	CGCCGGCCAC	GCTCGACGTA	AGCGGGGAGA	TCGCATCCCC	GTCCGGCTCT	CCGTGAGAG	5400
	GCCCCACGTG	CGCGTCCGGT	GCGGTCCCCG	CAATCGACTC	TATGGGCGTC	GTATTGCGCG	5460
	ACGCGCCAGG	ATCATGGTTC	TGGGGTGCAA	ACGTCCAGCC	CCACGAGGCA	AGGATAGTAA	5520
	ACGAGAGGG	GGACCTCTC	TTCCCCACG	CCCGACATCA	CGGACCGGTA	TGAGACCCGA	5580
25	GATTTAACCA	TCAACAGTCT	TTATTAATTG	CCCCCTCGTA	AATCAAGACC	CCGGATGTCTG	5640
	GCATCTTATA	CCGACCAGTC	GATAGGCATA	ATGTCCCGGG	TTTCGAGGTA	GCGATTGCGG	5700
	GCGAGGAAAT	GCTGGCACGT	CCCAAACGGG	ACCTTGAGAG	GGGGCGACGG	GTGAGAAAAC	5760
	TTGAGGACGT	AGTGTGCGG	AGGGTCGGGC	CTGATCGCGT	TCTGGGCATG	GGCGCCCCAG	5820
	AGCATAAAGA	CCAGGCCCGG	GCGGCGCGCG	GCCAGCCGTC	GGACCACCCC	GCCCACAAAG	5880
30	CGGTCCCATC	CAAGCTTGA	GTGGGACGCC	GCCGCCCCGC	GCTTGACGGT	CAGGGTCGTG	5940
	TTCAACAACA	GCACGCCGTC	GCGAGCCCAC	TTTTCCAGGC	AGCCGCGGCC	GCTCATGCGC	6000
	GCGTCGGGGT	AACAATTTTT	TACCGCCGCC	AGCACGTTCC	GTAGACTCGG	AGGCACCGGC	6060
	ACATCCGCAC	GCACGCTAAA	CGCCAGGCCG	TGCGCCTGGC	CGGGATGGTG	GTACGGGTCC	6120
	TGCCCGATGA	TAACCACGCG	CACGTCGTCG	GGGGTACAAT	ACCGCGTCCA	GGAGAACACA	6180
35	TCCTCCCGCG	GCGGCAGCAC	CTCTTCGGTC	TGGCACCGAC	GGTCATACTC	CGCGAGGAGG	6240
	CGCGCGGTTA	GGGGGTTTCG	GAGCTCCGGC	TCCAACAGGG	GCCGCCAGGC	GTCGTCGATC	6300
	AGAAACGCGC	GCCGAAACGC	GGCCCAGTCC	AGGGGCACGG	AAGTCGGCAG	GGGCGGCGTC	6360
	GCATCGCCCG	ACAGCCCCAG	GGGCTGTTTC	GGGGTTGTAG	ATGCGGAAAA	CATCACGCCC	6420
	GGCGGACAGC	CTCTAGGGCG	ACGACGCTGG	ACAGCCGACC	CGGCCGCTAG	ACGGGTCGGA	6480
40	TATCCTGCTC	CCGACCCAGG	GTGGCTTGCG	ATGCAGACGT	GGCTAGTTGC	GTCGGAGGCG	6540
	AGTATGCCCG	CGCCGACCTC	GCGTCGGGGA	GACCCGCCGT	GGGGGGGCGT	TCGAAAGGGC	6600
	GAGGACGGGC	GGCTGGGTGG	CGAGGGGCTT	CGACTGCGAG	CTCGCTTCAT	CGTCCGACGA	6660
	CACAGGCGGT	TCCCACGTTG	ACGTGGTGT	GGCAGGCCGT	AAATCGCGTC	GCCCGACGCA	6720

	GCGGCGAGTG	CGTGAGTAGT	CAAAGTTTAC	ACACCCGGCC	CTGACGGGTG	CGTGGCTGAC	6780
	GGCCTGTTTT	CGACTGCCCA	ACGCATCGCG	TATCTCTTTA	TAAAGGGCCC	GGCGCGTCGT	6840
	TGTTTCCTGG	GTGTCGGCCG	GAAACACAGA	GTGACTCAAG	TCCTCCAAAA	ATCCCCCCGC	6900
	AAAGAGAAAG	GGGTTAACCA	GATACGCCCT	CTGGGCGTGC	CTATCCCACA	AAAACGTGTC	6960
5	CAACCCCGGG	CAGTGATAGC	GAAGAAATAT	TCCGTCTATG	CGGGCATAGT	CAATAACGGA	7020
	CGGGGCTTCG	TAGCGCCAAG	AAACATCGTC	CGCGGGGGTC	CGCATGCAAG	GCACTCTTAG	7080
	TATGTCCCCC	ACCTCTTTGG	CAATAACACT	ACGAAGAACA	TATTCGGTTG	CCTGTGACCC	7140
	ACCCACAGCC	CCCCAGGGTC	CCATAACGAC	AAGCCCAAAC	AGACAGACGA	ACCCCATAGC	7200
	GAGCGGACAG	TGTAACCGGT	AAGCCCCCTT	GTTCCCGCAT	AAAAAACGTC	CAAACAAGAC	7260
10	AACCGCGAGC	AACCGAATCA	CGCGGGTCCA	ATATGCCCAT	TCCCGCGCTT	TCTACCGCTT	7320
	TATATATCCC	CCGTGTCCCT	CCCTCCCCCG	CGTCCCTCCA	TCCCCGCGT	CCTCCCTTCC	7380
	CCCGCGTCTT	CCCATCCCCC	GCGTCTCTCC	CTCCCCTGCG	CACACGTGAT	AGGTTTGGG	7440
	AACCCGAGGG	GCGACGCGGG	GAAAGCGCGC	CCCCGCCCGG	CCGCCGAGCG	CCCCGCCCGG	7500
	GCAGCCGAGC	GCCCCCGCCC	GGCCACCGCG	AGCCCCCGCC	CGGCCGCCCG	GGTCGCGCCG	7560
15	GCGCCCCCTC	CCGGCGCTTC	CGGGGCCTTT	CTGTCTGTTC	CCGCCGGGAC	CCCGGCCCGG	7620
	CCCCACCGCC	CCGCCCGGCA	GGGGGGCCCC	GGCGCCGCGC	AAAACACACA	GACGAACACA	7680
	CGGTGGCGAT	CTTTCTTTTA	CTTCGGCAGA	CCAGCGAGCC	CCGGCCCCGG	CCCGCGCCCC	7740
	GCGGCCACAC	CCACGGCACC	CCCCCGCCG	CCCACCCCGG	GGTCCACACA	GGAGCGCGCG	7800
	GGCGGCAAAA	ACGCGGGCGT	TTTTTTTTTT	TTTTTTTCCC	CTTTTCTCTC	TCTTTTCTCT	7860
20	CCCCTCTTTC	TTTTCTTTC	CCTTTTTCTT	CTTCCCTCTC	TCTTTCTTTC	TCCTTCTCTC	7920
	TCTTCTCTTT	CTCTTCTCT	TTCTTCTCT	CTCTTTCTTT	CTCCTCTCTC	TATCCTCTTA	7980
	TCTGTCCACT	TCTCCTTCTT	TTTCTTTGCC	TGCTGTTTCT	CTTTTCTTTT	CTCTTTCTTC	8040
	TATTCTCTGC	CTCTCTCCTT	CTTACTTCTC	TTCTCTCTCT	CTATCTACTG	TCACTCTATC	8100
	TCTTTCCCTT	CTTTTATATG	TGTCGTATTC	ATTCTTTCTC	TACTCACGTT	ACTCTATCTA	8160
25	TCTTCTTCAT	CATCTCTCTC	TCCTACTCTC	TCTCCCTCCT	TACTACTTCT	CTCTTCTCTT	8220
	TTTCTTTTAA	TATATTTTCT	TTCTTTACTC	ATCCCTTTTT	TCACTTTACT	ATTCCATGTG	8280
	TATCTCTTCT	CTCTTCTTTT	TTTCTCTCTC	TTTATTTCTC	ACTTTCTCCC	TCCTCCACTC	8340
	TTCTCATCTT	TTTTTCTCTA	CTCAACATAT	CTCTCATCTC	TCTCTCTCTA	TTTACTACTC	8400
	CTACTTCTTT	TTTCTCTCTC	TCATATTCTA	TTTTCACTTT	TCTTCTCTCT	TCTCATATCT	8460
30	CACCTCTCGT	TCTCTCCCTC	TTTTCATTTT	ATCTCTTATA	TCCTCTCTAT	CTTTATTCTT	8520
	CCTCTCTCAT	TCTCTCTCCT	CTGCTCACAC	TTACTTACTC	CTCCTCTCCA	ATTGTCTCTT	8580
	TCGTCTCTCA	CTCTTCTCCT	TTCTCATTTT	TTTCTATACT	CTCTCTATCT	TCGTATCTCT	8640
	TTTATCTATC	TACTTCTCTA	TTCTATCTCC	TCTATGCTTG	ACTCTCGTTA	CATTCTCATC	8700
	CCTTCGCTCC	TTATCAATTA	TCTTCGCACT	GCTTANGTAT	TCTCTCTCTC	TCTCCTCTCT	8760
35	CTCTCATTTT	CTCCTCTCCT	GCTTCTCTCT	CTTTTCTTGT	GGCTCTATCC	ACTTCTTCAT	8820
	TATATTCTCT	CACTTATTTT	CTTCTTTCTC	TCTCACTGCT	CTCTCACCTC	TCTCTCTACT	8880
	TTCTCTCTTC	CTCTGTTTTT	TCTCCTCTCT	TTGTCTATTC	ATCCCTCTTT	AACGTTCTTC	8940
	TTCTCTTCCG	TC					8953

## 40 (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 594 amino acids

(B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

10	Met	Val	Leu	Met	Gly	Arg	Leu	Arg	Asn	Ala	Pro	Glu	Ser	Leu	Thr	Tyr
	1				5					10					15	
	Met	Phe	Cys	Ala	Ala	Ile	Arg	Val	Ala	Pro	Val	Thr	Thr	Gln	Ser	Arg
				20					25					30		
	Thr	Ser	Leu	Arg	Val	Cys	Thr	His	Val	Leu	Phe	Pro	Asp	Pro	Ala	Leu
			35					40					45			
15	Pro	Val	Met	Arg	Tyr	Ala	Ala	Asn	Gly	Asn	Ser	Arg	Ser	Gly	Arg	Pro
		50					55				60					
	Val	Gly	Thr	Ser	Lys	Ala	Ala	Thr	Ser	Arg	Asn	His	Cys	Arg	Arg	Gly
	65				70					75				80		
	Thr	Cys	Val	Thr	Ser	Ser	Cys	Cys	Cys	Glu	Ser	Ser	Arg	Met	Arg	Ala
20					85					90				95		
	Met	Ile	Gly	Trp	Thr	Pro	Cys	Met	Asp	Val	Lys	Phe	Lys	Asn	Ala	Ser
				100				105					110			
	Ser	Leu	Asn	Arg	Thr	Ala	Gly	Leu	Ala	Pro	Gly	Cys	Cys	Gly	Gly	Gly
			115				120					125				
25	Pro	Gly	Ala	Arg	Thr	Ser	Arg	Glu	Pro	Ser	Pro	Pro	Asp	Ala	Ala	Met
		130					135					140				
	Ala	Ala	Gln	Arg	Ala	Arg	Ala	Pro	Ala	Met	Arg	Thr	Arg	Gly	Gly	Asp
	145				150					155				160		
	Ala	Ala	Leu	Cys	Ala	Pro	Glu	Asp	Gly	Trp	Val	Lys	Val	His	Pro	Thr
30				165					170					175		
	Pro	Gly	Thr	Met	Leu	Phe	Arg	Glu	Ile	Leu	Leu	Gly	Gln	Met	Gly	Tyr
				180				185					190			
	Thr	Glu	Gly	Gln	Gly	Val	Tyr	Asn	Val	Val	Arg	Ser	Ser	Glu	Ala	Ala
		195					200					205				
35	Thr	Arg	Gln	Leu	Gln	Ala	Ala	Ile	Phe	His	Ala	Leu	Leu	Asn	Ala	Thr
		210				215					220					
	Tyr	Asp	Leu	Glu	Glu	Asp	Trp	Arg	Arg	His	Val	Val	Arg	Leu	Gln	Pro
	225				230					235				240		
	Gln	Arg	Leu	Val	Arg	Arg	Tyr	Arg	Asn	Ala	Arg	Glu	Gly	Asp	Ile	Ala
40				245						250				255		
	Gly	Val	Ala	Glu	Arg	Val	Phe	Asp	Thr	Trp	Arg	Cys	Thr	Leu	Arg	Thr
			260				265					270				
	Thr	Leu	Leu	Asp	Phe	Ala	His	Gly	Val	Val	Asn	Cys	Phe	Ala	Pro	Gly

	275		280		285
	Gly Pro Ser Gly Pro Thr Ser Phe Pro Lys Tyr Ile Asp Trp Leu Thr				
	290		295		300
5	Cys Leu Gly Leu Val Pro Ile Leu Arg Lys Thr Arg Glu Gly Glu Ala				
	305		310		315
	Thr Gln Arg Leu Gly Ala Phe Leu Arg Gln His Thr Leu Pro Arg Gln				
		325		330	
	Leu Ala Thr Val Ala Gly Ala Ala Glu Arg Ala Gly Pro Gly Leu Leu				
		340		345	
10	Glu Leu Ala Val Ala Phe Asp Ser Thr Arg Met Ala Glu Tyr Asp Arg				
		355		360	
	Val His Ile Tyr Tyr Asn His Arg Arg Gly Glu Trp Leu Val Arg Asp				
		370		375	
	Pro Val Ser Gly Gln Arg Gly Glu Cys Leu Val Leu Cys Pro Pro Leu				
15			390		395
	Trp Thr Gly Asp Arg Leu Val Phe Asp Ser Pro Val Gln Arg Leu Cys				
		405		410	
	Pro Glu Ile Val Ala Cys His Ala Leu Arg Glu His Ala His Ile Cys				
		420		425	
20	Arg Leu Arg Asn Thr Ala Ser Val Lys Val Leu Leu Gly Arg Lys Ser				
		435		440	
	Asp Ser Gly Val Ala Gly Ala Ala Arg Val Val Asn Lys Ala Leu Gly				
		450		455	
	Glu Asp Asp Glu Thr Lys Ala Gly Ser Ala Ala Ser Cys Leu Val Arg				
25			470		475
	Leu Ile Ile Asn Met Lys Gly Met Arg His Val Gly Asp Ile Asn Asp				
		485		490	
	Thr Val Arg Ala Tyr Leu Asp Glu Ala Gly Gly His Leu Ile Asp Thr				
		500		505	
30	Pro Ala Val Asp His Thr Leu Pro Gly Phe Gly Lys Gly Gly Thr Gly				
		515		520	
	Arg Gly Ser Ala Ala Gln Asp Pro Gly Ala Arg Pro Gln Gln Leu Arg				
		530		535	
	Gln Ala Phe Gln Thr Ala Val Val Asn Asn Ile Asn Gly Met Leu Glu				
35			550		555
	Gly Tyr Ile Asn Asn Leu Phe Gly Thr Ile Glu Arg Leu Arg Glu Thr				
		565		570	
	Asn Ala Gly Leu Ala Thr Gln Leu Gln Arg Gly Ser Ser Arg Ser Thr				
		580		585	
40	Ala Xaa				590

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 877 amino acids

(B) TYPE: amino acid

5 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Ala Ala Ser Gly Gly Glu Gly Ser Arg Asp Val Arg Ala Pro Gly  
 1 5 10 15  
 Pro Pro Pro Gln Gln Pro Gly Ala Arg Pro Ala Val Arg Phe Arg Asp  
 15 20 25 30  
 Glu Ala Phe Leu Asn Phe Thr Ser Met His Gly Val Gln Pro Ile Ile  
 35 40 45  
 Ala Arg Ile Arg Glu Leu Ser Gln Gln Gln Leu Asp Val Thr Gln Val  
 50 55 60  
 20 Pro Arg Leu Gln Trp Phe Arg Asp Val Ala Ala Leu Glu Val Pro Thr  
 65 70 75 80  
 Gly Leu Pro Leu Arg Glu Phe Pro Phe Ala Ala Tyr Leu Ile Thr Gly  
 85 90 95  
 Asn Ala Gly Ser Gly Lys Ser Thr Cys Val Gln Thr Leu Asn Glu Val  
 25 100 105 110  
 Leu Asp Cys Val Val Thr Gly Ala Thr Arg Ile Ala Ala Gln Asn Met  
 115 120 125  
 Tyr Val Lys Leu Ser Gly Ala Phe Leu Ser Arg Pro Ile Asn Thr Ile  
 130 135 140  
 30 Phe His Glu Phe Gly Phe Arg Gly Asn His Val Gln Ala Gln Leu Gly  
 145 150 155 160  
 Gln His Pro Tyr Thr Leu Ala Ser Ser Pro Ala Ser Leu Glu Asp Leu  
 165 170 175  
 Gln Arg Arg Asp Leu Thr Tyr Tyr Trp Glu Val Ile Leu Asp Ile Thr  
 35 180 185 190  
 Lys Arg Ala Ala His Gly Gly Glu Asp Ala Arg Asn Glu Phe His Ala  
 195 200 205  
 Leu Thr Ala Leu Glu Gln Thr Leu Gly Leu Gly Gln Gly Ala Leu Thr  
 210 215 220  
 40 Arg Leu Ala Ser Val Thr His Gly Ala Leu Pro Ala Phe Thr Arg Ser  
 225 230 235 240  
 Asn Ile Ile Val Ile Asp Glu Ala Gly Leu Leu Gly Arg His Leu Leu  
 245 250 255



Thr Thr Val Val Tyr Cys Trp Trp Met Ile Asn Ala Leu Tyr His Thr  
 260 265 270  
 Pro Gln Tyr Ala Gly Arg Leu Arg Pro Val Leu Val Cys Val Gly Ser  
 275 280 285  
 5 Pro Thr Gln Thr Ala Ser Leu Glu Ser Thr Phe Glu His Gln Lys Leu  
 290 295 300  
 Arg Cys Ser Val Arg Gln Ser Glu Asn Val Leu Thr Tyr Leu Ile Cys  
 305 310 315 320  
 Asn Arg Thr Leu Arg Glu Tyr Thr Arg Leu Ser His Ser Trp Ala Ile  
 10 325 330 335  
 Phe Ile Asn Asn Lys Arg Cys Val Glu His Glu Phe Gly Asn Leu Met  
 340 345 350  
 Lys Val Leu Glu Tyr Gly Leu Pro Ile Thr Glu Glu His Met Gln Phe  
 355 360 365  
 15 Val Asp Arg Phe Val Val Pro Glu Ser Tyr Ile Thr Asn Pro Ala Asn  
 370 375 380  
 Leu Pro Gly Trp Thr Arg Leu Phe Ser Ser His Lys Glu Val Ser Ala  
 385 390 395 400  
 Tyr Met Ala Lys Leu His Ala Tyr Leu Lys Val Thr Arg Glu Gly Glu  
 20 405 410 415  
 Phe Val Val Phe Thr Leu Pro Val Leu Thr Phe Val Ser Val Lys Glu  
 420 425 430  
 Phe Asp Lys Tyr Arg Arg Leu Thr Gln Gln Pro Thr Leu Thr Met Glu  
 435 440 445  
 25 Lys Trp Ile Thr Ala Asn Ala Ser Arg Ile Thr Asn Tyr Ser Gln Ser  
 450 455 460  
 Gln Asp Gln Asp Ala Gly His Val Arg Cys Glu Val His Ser Lys Gln  
 465 470 475 480  
 Gln Leu Val Val Ala Arg Asn Asp Ile Thr Tyr Val Leu Asn Ser Gln  
 30 485 490 495  
 Val Ala Val Thr Ala Arg Leu Arg Lys Met Val Phe Gly Phe Asp Gly  
 500 505 510  
 Thr Phe Arg Thr Phe Glu Ala Val Leu Arg Asp Asp Ser Phe Val Lys  
 515 520 525  
 35 Thr Gln Gly Glu Thr Ser Val Glu Phe Ala Tyr Arg Phe Leu Ser Arg  
 530 535 540  
 Leu Met Phe Gly Gly Leu Ile His Phe Tyr Asn Phe Leu Gln Arg Pro  
 545 550 555 560  
 Gly Leu Asp Ala Thr Gln Arg Thr Leu Ala Tyr Gly Arg Leu Gly Glu  
 40 565 570 575  
 Leu Thr Ala Glu Leu Leu Ser Leu Arg Arg Asp Ala Ala Gly Ala Ser  
 580 585 590  
 Ala Thr Arg Ala Ala Asp Thr Ser Asp Arg Ser Pro Gly Glu Arg Ala

	595		600		605
	Phe Asn Phe Lys His Leu Gly	Pro Arg Asp Gly Gly	Pro Asp Asp Phe		
	610	615	620		
5	Pro Asp Asp Asp Leu Asp Val Ile Phe Ala Gly Leu Asp Glu Gln Gln				
	625	630	635	640	
	Leu Asp Val Phe Tyr Cys His Tyr Ala Leu Glu Glu Pro Glu Thr Thr				
	645	650	655		
	Ala Ala Val His Ala Gln Phe Gly Leu Leu Lys Arg Ala Phe Leu Gly				
	660	665	670		
10	Arg Tyr Leu Ile Leu Arg Glu Leu Phe Gly Glu Val Phe Glu Ser Ala				
	675	680	685		
	Pro Phe Ser Thr Tyr Val Asp Asn Val Ile Phe Arg Gly Cys Glu Leu				
	690	695	700		
	Leu Thr Gly Ser Pro Arg Gly Gly Leu Met Ser Val Gln Thr Asp Asn				
15	705	710	715	720	
	Tyr Thr Leu Met Gly Tyr Thr Tyr Thr Arg Val Phe Ala Phe Ala Glu				
	725	730	735		
	Glu Leu Arg Arg Arg His Ala Thr Ala Gly Val Ala Glu Phe Leu Glu				
	740	745	750		
20	Glu Ser Pro Leu Pro Tyr Ile Val Leu Arg Asp Gln His Gly Phe Met				
	755	760	765		
	Ser Val Val Asn Thr Asn Ile Ser Glu Phe Val Glu Ser Ile Asp Ser				
	770	775	780		
	Thr Glu Leu Ala Met Ala Ile Asn Ala Asp Tyr Gly Ile Ser Ser Lys				
25	785	790	795	800	
	Leu Ala Met Thr Ile Thr Arg Ser Gln Gly Leu Ser Leu Asp Lys Val				
	805	810	815		
	Ala Ile Cys Phe Thr Pro Gly Asn Leu Arg Leu Asn Ser Ala Tyr Val				
	820	825	830		
30	Ala Met Ser Arg Thr Thr Ser Ser Glu Phe Leu His Met Asn Leu Asn				
	835	840	845		
	Pro Leu Arg Glu Arg His Glu Arg Asp Asp Val Ile Ser Glu His Ile				
	850	855	860		
	Leu Ser Ala Leu Arg Asp Pro Asn Val Val Ile Val Tyr				
35	865	870	875		

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 199 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

5  
Met Gly Asn Pro Gln Thr Thr Ile Ala Tyr Ser Leu His His Pro Arg  
1 5 10 15  
Ala Ser Leu Thr Ser Ala Leu Pro Asp Ala Ala Gln Val Val His Val  
20 25 30  
10 Phe Glu Ser Gly Thr Arg Ala Val Leu Thr Arg Gly Arg Ala Arg Gln  
35 40 45  
Asp Arg Leu Pro Arg Gly Gly Val Val Ile Gln His Thr Pro Ile Gly  
50 55 60  
Leu Leu Val Ile Ile Asp Cys Arg Ala Glu Phe Cys Ala Tyr Arg Phe  
15 65 70 75 80  
Ile Gly Arg Ala Ser Thr Gln Arg Leu Glu Arg Trp Trp Asp Ala His  
85 90 95  
Met Tyr Ala Tyr Pro Phe Asp Ser Trp Val Ser Ser Ser His Gly Glu  
100 105 110  
20 Ser Val Arg Ser Ala Thr Ala Gly Ile Leu Thr Val Val Trp Thr Pro  
115 120 125  
Asp Thr Ile Tyr Ile Thr Ala Thr Ile Tyr Gly Thr Ala Pro Glu Ala  
130 135 140  
Arg Cys Asp Asn Ala Pro Leu Asp Val Arg Pro Thr Thr Pro Pro Ala  
25 145 150 155 160  
Pro Val Ser Pro Thr Ala Gly Glu Phe Pro Ala Asn Thr Thr Asp Leu  
165 170 175  
Leu Val Glu Val Leu Arg Glu Ile Gln Ile Ser Pro Thr Leu Asp Asp  
180 185 190  
30 Ala Asp Pro Thr Pro Gly Thr  
195

(2) INFORMATION FOR SEQ ID NO:5:

35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 172 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Val Gly Pro Leu Asp Gly Glu Pro Asp Arg Asp Ala Ile Ser Pro Leu  
 1 5 10 15  
 Thr Ser Ser Val Ala Gly Asp Pro Pro Gly Ala Asp Gly Pro Tyr Val  
 5 20 25 30  
 Thr Phe Asp Thr Leu Phe Met Val Ser Ser Ile Asp Glu Leu Gly Arg  
 35 40 45  
 Arg Gln Leu Thr Asp Thr Ile Arg Lys Asp Leu Arg Leu Ser Leu Ala  
 50 55 60  
 10 Lys Phe Ser Ile Ala Cys Thr Lys Thr Ser Ser Phe Ser Gly Thr Ala  
 65 70 75 80  
 Ala Arg Gln Arg Lys Arg Gly Ala Pro Pro Gln Arg Thr Cys Val Pro  
 85 90 95  
 Arg Ser Asn Lys Ser Leu Gln Met Phe Val Leu Cys Lys Arg Ala Asn  
 15 100 105 110  
 Ala Ala Gln Val Arg Glu Gln Leu Arg Ala Val Ile Arg Ser Arg Lys  
 115 120 125  
 Pro Arg Lys Tyr Tyr Thr Arg Ser Ser Asp Gly Arg Leu Cys Pro Ala  
 130 135 140  
 20 Val Pro Val Phe Val His Glu Phe Val Ser Ser Glu Pro Met Arg Leu  
 145 150 155 160  
 His Arg Asp Asn Val Met Leu Ser Thr Glu Pro Asp  
 165 170

25 (2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 334 amino acids  
 (B) TYPE: amino acid  
 30 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Lys Arg Ala Arg Ser Arg Ser Pro Ser Pro Pro Ser Arg Pro Ser  
 1 5 10 15  
 Ser Pro Phe Arg Thr Pro Pro His Gly Gly Ser Pro Arg Arg Glu Val  
 40 20 25 30  
 Gly Ala Gly Ile Leu Ala Ser Asp Ala Thr Ser His Val Cys Ile Ala  
 35 40 45  
 Ser His Pro Gly Ser Gly Ala Gly Tyr Pro Thr Arg Leu Ala Ala Gly

	50		55		60											
	Ser	Ala	Val	Gln	Arg	Arg	Arg	Pro	Arg	Gly	Cys	Pro	Pro	Gly	Val	Met
	65				70					75						80
	Phe	Ser	Ala	Ser	Thr	Thr	Pro	Glu	Gln	Pro	Leu	Gly	Leu	Ser	Gly	Asp
5					85					90					95	
	Ala	Thr	Pro	Pro	Leu	Pro	Thr	Ser	Val	Pro	Leu	Asp	Trp	Ala	Ala	Phe
					100					105					110	
	Arg	Arg	Ala	Phe	Leu	Ile	Asp	Asp	Ala	Trp	Arg	Pro	Leu	Leu	Glu	Pro
					115					120					125	
10	Glu	Leu	Ala	Asn	Pro	Leu	Thr	Ala	Arg	Leu	Leu	Ala	Glu	Tyr	Asp	Arg
					130					135					140	
	Arg	Cys	Gln	Thr	Glu	Glu	Val	Leu	Pro	Pro	Arg	Glu	Asp	Val	Phe	Ser
	145											155				160
	Trp	Thr	Arg	Tyr	Cys	Thr	Pro	Asp	Asp	Val	Arg	Val	Val	Ile	Ile	Gly
15					165						170					175
	Gln	Asp	Pro	Tyr	His	His	Pro	Gly	Gln	Ala	His	Gly	Leu	Ala	Phe	Ser
					180						185				190	
	Val	Arg	Ala	Asp	Val	Pro	Val	Pro	Pro	Ser	Leu	Arg	Asn	Val	Leu	Ala
					195						200				205	
20	Ala	Val	Lys	Asn	Cys	Tyr	Pro	Asp	Ala	Arg	Met	Ser	Gly	Arg	Gly	Cys
					210						215				220	
	Leu	Glu	Lys	Trp	Ala	Arg	Asp	Gly	Val	Leu	Leu	Leu	Asn	Thr	Thr	Leu
	225												235			240
	Thr	Val	Lys	Arg	Gly	Ala	Ala	Ala	Ser	His	Ser	Lys	Leu	Gly	Trp	Asp
25					245						250					255
	Arg	Phe	Val	Gly	Gly	Val	Val	Arg	Arg	Leu	Ala	Ala	Arg	Arg	Pro	Gly
					260						265					270
	Leu	Val	Phe	Met	Leu	Trp	Gly	Ala	His	Ala	Gln	Asn	Ala	Ile	Arg	Pro
					275						280				285	
30	Asp	Pro	Arg	Gln	His	Tyr	Val	Leu	Lys	Phe	Ser	His	Pro	Ser	Pro	Leu
					290						295				300	
	Ser	Lys	Val	Pro	Phe	Gly	Thr	Cys	Gln	His	Phe	Leu	Ala	Ala	Asn	Arg
	305										310				315	320
	Tyr	Leu	Glu	Thr	Arg	Asp	Ile	Met	Pro	Ile	Asp	Trp	Ser	Val		
35					325						330					

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 183 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

5  
Val Pro Cys Met Arg Thr Pro Ala Asp Asp Val Ser Trp Arg Tyr Glu  
1 5 10 15  
Ala Pro Ser Val Ile Asp Tyr Ala Arg Ile Asp Gly Ile Phe Leu Arg  
20 25 30  
10 Tyr His Cys Pro Gly Leu Asp Thr Phe Leu Trp Asp Arg His Ala Gln  
35 40 45  
Arg Ala Tyr Leu Val Asn Pro Phe Leu Phe Ala Gly Gly Phe Leu Glu  
50 55 60  
Asp Leu Ser His Ser Val Phe Pro Ala Asp Thr Gln Glu Thr Thr Thr  
15 65 70 75 80  
Arg Arg Ala Leu Tyr Lys Glu Ile Arg Asp Ala Leu Gly Ser Arg Lys  
85 90 95  
Gln Ala Val Ser His Ala Pro Val Arg Ala Gly Cys Val Asn Phe Asp  
100 105 110  
20 Tyr Ser Arg Thr Arg Arg Cys Val Gly Arg Arg Asp Leu Arg Pro Ala  
115 120 125  
Asn Thr Thr Ser Thr Trp Glu Pro Pro Val Ser Ser Asp Asp Glu Ala  
130 135 140  
Ser Ser Gln Ser Lys Pro Leu Ala Thr Gln Pro Pro Val Leu Ala Leu  
25 145 150 155 160  
Ser Asn Ala Pro Pro Arg Arg Val Ser Pro Thr Arg Gly Arg Arg Arg  
165 170 175  
His Thr Arg Leu Arg Arg Asn  
180

30

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 9218 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CCTCCGGACG TGCGATCGGA TCCCGCGAGT CGAAATCCCA CACAGCAGAC CCGTGGGTGT 60  
GCTAGATCGA ACGAGCGGCA GGATCGCGTG CTGGCCCCCTT GATACGATCT CGTCGACCGG 120

	GGACTCCCTT	CTACCCCCAC	CCAACCAGCG	CGCCGGCGCT	TAGGGTGTGA	CCCCCCCCAT	180
	GGCATCCGGG	GTTCCTCCCG	CCCACCCCCA	AACCCCGGTT	GGGGCGGGCA	GCCGAGACCT	240
	TTCTTTAAAA	GGCACCCCAT	CCGACGGCAT	GCAGCCCAGA	GGAGCGGACA	CGCTTGAAGG	300
	GCACTCGCTT	CCGACCGACG	GGCCCCCGCA	CCGGGGCGGC	GACCATGATC	CGGCGGCGGG	360
5	GAAACGTGGA	GATTTCGGTC	TACTACGAGT	CTGTGCGGCC	CTCTCGATCC	CGAAGCCATC	420
	TGAAGCCGTC	CGACCATCAA	GAATTCCCAG	GGCACCACGT	GTCCCCAGGG	AGCCCCGGGT	480
	TCCCCGAGAG	CCCAGGGAAC	CGCGAGTTCC	ACGATCTCCC	AGAGAACCCA	GGGTCCCGCG	540
	CATACCCAGG	GACCCGCGAC	CCCCACGACC	CCCACGGGTG	CCCAGGGAGC	CTAGACCCCC	600
	ACGGGAACCC	CGCGCAACCC	GCGGGCTTGC	CTAGCCCGGT	CCCCCTACGC	CCCCCTCGGCA	660
10	GCCCCGACCC	CTCATCGCCG	CGCCAACGCA	CGTACGTTCT	GCCCCGCGTC	GGGATCCGTA	720
	ACGCGCCCCG	GTCCGACACC	CGGGCCCCAA	AGCGTGCCCA	CTCGCGGCAC	CGCGCGGACC	780
	GGCCCCCGGA	GTCCCCCGGC	TCCGAGTTGT	ACCTCTCAA	CGCCCAGGCC	CTGGCGCACC	840
	TGCAGATGCT	GCCCCGCGAC	CACCGGGCCT	TTTTTCGGAC	GGTGATCGAG	GTGTCCCGCC	900
	TGTGTGCTCT	CAACACCCAC	GACCCACCGC	CCCCGCTGGC	GGGAGCCAGG	GTCCGACAGG	960
15	AGGCGCAGCT	GGTTCATACC	CAATGGCTTC	GGGCCAACAG	GGAGTCCTCG	CCGCTGTGGC	1020
	CCTGGCGGAC	GGCCGCCATG	AATTTTATCG	CCGCGGCTGC	GCCCTGCGTC	CAAACACATC	1080
	GCCATATGCA	CGACCTGCTG	ATGGCATGCG	CCTTCTGGTG	CTGTTTGGCG	CACGCGTCGA	1140
	CGTGTTCTTA	CGCGGGGTTA	TATTCGGCAC	ACTGCCAGCA	TTTGTTTCGT	GCGTTTGGGT	1200
	GCGGACCCCC	GGTCCTGACC	ACGTCCCGGG	GACAGGGTGG	TTGGTGTAAT	TAATAATAAA	1260
20	ATCGTGAAAA	TTGAAATCGC	TTTGTGTGTT	GCTGCGGGGA	CGGGGGCAAA	TGCGTCGTGA	1320
	CTCTAGAACG	CCAGATGTGG	GGTGCGGATG	GGGAAATGTA	TGGGTCTTTC	GTCTGGAGCC	1380
	CGTACCCGGC	AGAGAGATTT	CCCCAGCACG	GAGGAACTGG	GGTACTGCAC	TGCCCCCTC	1440
	CTGGGGGGGG	GGGGGGCGAG	AGGTCAATAG	ATTTCCCCAA	GAGACTTCCC	TAACACGGAG	1500
	GAGCCGGGAG	AGTTCAATAG	ATTTCCCAAA	CAC TGAGGAA	CTGGGGTACT	GCACTGCCCC	1560
25	CTCTGGGGG	GGGGGGCGAG	AGGTCAATAG	ATTTCCCAAA	GAGACTTCCC	TAACACGGAG	1620
	GAGCCGGGAG	AGTTCAATAG	ATTTCCCAAA	CAC TGAGGAA	CTGGGGTACT	GCACTTCCCC	1680
	CCCCCGGGGG	GTGAAATTCC	GAGAATTTTT	TACCCTTTTT	TGCATTTCTT	TCCCCCCCCC	1740
	CCCCAAAAAA	AAAGACAACC	TAGTAGACCG	TAATGACAAT	CAACCACTTT	ATTGCAATTA	1800
	ACATACGGAC	GTGGGTGCGG	GCGAGGGGTG	GGGGCGAAGA	AGGCGCCATA	CATCGAGGCG	1860
30	TCATTTAGCG	GAGCAGCCAC	ACCAAAAGTG	CCCCGAACCC	TCCAGATAGG	AGGGCCACGA	1920
	CGAGACAGGC	GATAACCAGC	CCGACGCACC	GCGTGCGCCG	CCGTGCGCGC	CTTAGGACCG	1980
	ACTGCTGGCG	GCCCATGCGC	ACGAGGAAGT	CGTTGGCGGC	CTCGTCTTCG	CTTTCCGAGT	2040
	AGTAGGCTTC	TGCCGGGACG	GGCGAGGCCG	CGGGGTAAAG	CGGCACCGAC	GCGCTGGAAC	2100
	GCACCGAGTC	TTGGTCGGCG	GGCCGGGAGG	TCATCGCGGA	CGCGGAAGGG	CGCTGGCGGA	2160
35	GGGCCGGAGG	CGAAGGTGCG	GTTGCCGTGA	CTCACGATTT	TTATGAGCTG	CGGCGGGGCT	2220
	GGCCGCCGGA	CCTTTATGCG	CCTCGGGCGA	TTGACGTCAC	GTAAAACGCA	ATCCCCCACA	2280
	GGACGGCCCC	GAGACCCACC	GCCCCCGCA	GCCAGCGCAC	GGCGAGCCAG	GTGACGAATT	2340
	GGGAGGGGGC	GTCCACGGCG	TGGAGGGCCA	CGGGAAGGC	CGCGGGGGAG	CCGCCGCGAG	2400
	GTGGTCTGCG	GCACGCGGGC	GCGGCGCCGC	CCGCGCCGGG	GGGAGGGGTC	TCTGGCGGGT	2460
40	CCCCGCGTGC	GTCCGCGATG	GCAATCAGTT	CATCGCCGAC	GTCCGCGTCG	TCGGAAGACG	2520
	CCTTACCAGA	GGACGGACGA	ATAGGAGGCC	TGGGAGTGAC	GACGGCCCGG	GCTTCCCGAA	2580
	CCAAAGGTGG	TGAGCGGGCG	GCGAGATTTA	CGCCCCCTCG	TATGGGGGTA	TACAGACGGA	2640
	GCCGTTGGTG	ATAAGATCTC	AAAGCCGGAT	CCATTTGTGG	AGGGAGAGTC	GGGTCTCTCC	2700

	GGAGGGTCCT	GCCACAGGGA	CCCGTCGCGC	TCCCCCTCGC	TGTCCGAACT	CCAGTCCGCG	2760
	TACAGCTCGC	TGTCCGCCAC	GCGAATGTAA	GTGGGGCCCC	TCGCCGAGGC	CCGGCTTTTA	2820
	ACCGCCCCGC	AGGAGCGCCT	GCGCCAGCAG	GTCATGCACG	CCCACGCGGA	CAGCCCAGAG	2880
	GCGGCCAGCA	ACAGGGCCGC	CCCCAGCACC	GCCCCGAGGC	GCAGCGGGCC	GCGCGCGGAG	2940
5	GGCGCGGGAG	GGGGGGCTCT	CACGTGCGGG	CGGGTGGGCT	CGACGGGCTC	GGGCTGGCGC	3000
	TGGGGGAGGT	GCTGTTCAC	CACCGCGTTC	CGGTACTGCG	CCGCGGTGCT	GATGGTCATG	3060
	TGGCCCCAGG	CGTGGATATG	ATCGTCCACG	TACACCACGC	ACAGGTAGAG	GCCGGCGTGC	3120
	TGGGGGGAGG	CGTGCTGGAA	TTCCAGATTG	ACGGTGGAGG	CCAGCCACGC	CAACCCCGGG	3180
	ACCGGTTCCA	TGCGAGCCTC	GGCAAAACAT	CGCGGCGGGG	GCGTAGTCCT	GGAACAGCCG	3240
10	GCGTAGCTGC	GGACCGCCAG	GCGGTACGCC	CAGGAACCTTA	CGGCGCACGG	GCGTTCGGCC	3300
	GGAGATAGAC	ACTCTGGAAG	CTGCGGGTGA	TACAGACAAG	CTTCGTAGAT	CCGCATCTCG	3360
	GCGCACGAGG	ACGGCACGTC	AAACCGCATC	CAGACGACGT	CCATGGCGTA	CGGACCGTCG	3420
	TCGTGGGCTA	CGGCGTGGAT	GGAGACCTTC	GTCTCAAACG	TCTCCCCGGG	GGCAAACATA	3480
	ATGGCCTCCG	GGGTCTCCAT	ATGGACCGTT	ACCCCGCGCA	CGTGGGACAC	CTCGGGGATA	3540
15	ACACGATGGT	GCCTCGGGGG	GGCCACGGGG	GGACCAACGG	GGGGGGGTTG	GGGGGGGAAC	3600
	GCTGACCGGC	GTGCGTTTCG	TCACGCCCCG	GTCGTCTTCT	TCGTCGTAGT	CGTCGGGGGT	3660
	CGGGGTGCGC	ACAGGGGCGG	GCTCCACGAC	CAGAACCACC	GACGCCACTT	GGCGCGCCTC	3720
	GTCGCTTAGG	CCGACCACGG	ACAGGGTGTG	CAGACCGCTG	TCCGTCTCCA	GGGCCCCGTA	3780
	GATGACCAGA	CTCTCGTTGA	CCACGGCTAC	GCGATCGCGC	CACGCCAACT	CCGAATACAG	3840
20	TCCCTCGTCG	CCCGCGGGGA	ACGGGGGACT	GTATGCTATG	GCGAGCGGTT	CCGGGGCGCG	3900
	CATGCACGCC	GCATCCACGA	CCGTCTCGAG	CACCCGTCGG	GGGGGCCACA	GCGCCACCCA	3960
	CGACGGGCGC	AGGGGACCGC	AGGCATCCAG	GGGTTCGCG	GCCCACAGTA	GTTTGTGGGC	4020
	CCGGGTGCGT	TCCTCCGGCC	CCCGGGGCGC	CGGAAGCAAC	ACCACGTCTT	CGCCCGAGGT	4080
	TACCCGTTTC	CAGGACGTTT	TGGGTGCTGC	CGCCAGGCAC	GATACGACCC	AAACTCCAAC	4140
25	AAAAAACACC	AACCCGGCCC	CGCGAGCCAT	GTTGCGGTGG	CAGGAGCCGT	CGGTGCGGGC	4200
	AGATCGGAGA	CTAGCTGACG	GCGGCGCACC	AAGTCACCCG	AAGACACAGA	GTCGGGGCGG	4260
	CGACTCCTTA	AATGCGCGGC	GGGCCTCTCC	GACACTACCC	CCTTTATTCT	TTTTCCTCCC	4320
	CCCCCGGGCC	CCGCCATCC	ATTACCCGCC	TCCCATGCCA	TCCGGGGAAT	GACGAACGAT	4380
	CACAAAGGGA	TCCAACACAC	GCATATAGGC	AAATAACATC	GGTTTATTGG	GGGGGAAATA	4440
30	ACCACGATGG	GGGCGGTGGG	GCGGGCCTGC	CGAACGGCCC	GCTTGGACCT	AAACCTCTTG	4500
	GGGGGCCGTC	GGGCCACTGC	GGGGCCGAGG	ACTGACGGAC	AGCAGCACGA	CTGGACCTGG	4560
	CTCCGATTCC	TCAGCTATCG	ACGTTAGGGA	AGGCATGGTC	GTGGACGACG	ACGAACGGCG	4620
	TCGGGGTTTG	GGGGGGGTGT	TTGGGTGGGA	TCGCAGCTCG	GCTCCGAGGC	GGGCCATGGC	4680
	CGCCTCGTTG	ACCGCGCAGG	AAACGCCCCC	GGGGTTGTAA	ATCTGGCCGC	GGGGGCGCCT	4740
35	GTATCGGCGC	TGGCATCTAT	GGATGAAGCA	GATACAGCTG	CCCAGAAACA	CAAAGCGCAT	4800
	GATGGACGCC	GGTATGGCGA	TCTGGATTAC	CTGGGCTACG	GTTAGCCTGT	GTCTCGATTTC	4860
	GCTGGCCGAT	CGCGTGGAAT	TGGGCGGGGC	TCTCTCGCCG	CTCGCGGGCG	CGGGCGTCCC	4920
	TGTGTCCCCG	GGGGCGGGGG	TCGGGTCTCG	GGGGGAGGAC	GGGGATGTGC	TTGTCCGTGG	4980
	AGGGGTGGGC	CGGGAGGCTC	CGGGGTGTGA	TACGCTCGAG	GGTCCCAGGC	GCGGGGCCGA	5040
40	AAAGGGAAGC	TGCGCCGGAT	CGCAGGAGCC	GTAGTCCGAG	CCGTTATACA	CAAACGTCCC	5100
	GTTGGCAGAG	AGCGCCACCC	CCAAAACAAA	CAGGCTGGCG	TTCGTCGCGC	TGCCGACCCA	5160
	TACGCGCAGG	ACATACAGAC	CGGCATAGTC	GCGCGTTGCC	GTTCGAACCC	GCAGAAGCGG	5220
	CTGCCGCGCC	AGACCCAGCT	CCAGGTCGCG	ATAGGCGGGG	CTGTGGGCGT	GGTGCGTCGA	5280



	CGCACACAAG	GTGAACGCCA	CGGCGGGGCG	GCGGGGGCAT	GCGGTCAGTG	TGACCACGTG	5340
	TACAACGCGG	GGGCAGTGGT	TCCCCAGGGG	GTAAGTGAAC	AGCTCGATGA	TGCCGTCGTA	5400
	GTAAGTTGTG	TGGGGGACCT	GGGCCCCCAC	AAAATGAAGC	TCCCCGAAAA	CACGCAGGTC	5460
	CTCTTCCACG	AAGCCCTGGG	GCCCCACGGC	CCCGGCATCC	ACGAGTGAGT	CTGAGACCAG	5520
5	ACTGACCGTG	GGGCCGCGGA	CGACCAGGCC	GGTGGCGCAG	ACCCACAGGC	CCAGGATCGC	5580
	CAGGCCCTGC	AGCGAGCGGC	CGGGCATAAC	GGGATCGGAC	GGGTCGAGGT	GACTGTGGGC	5640
	GCGGTGGCTA	ATCGTCGGGA	CAGCGGTGCG	CGCCCCACGC	TCCCGGCCTA	TGAACTGTCC	5700
	TAGTTTCCCT	CCTTCGAGAC	TCCCTTTATG	CGGAGTCCAA	GTCCCACCCA	AATACCCAG	5760
	CCACCTCCC	ACACGGGCCC	AGAGGTACAC	GGGAGCGGGG	ATACTCCTCT	AGTAAACAA	5820
10	TGGCTGGTGC	GAGGGGGGCG	CGTCGTCATC	CCGGATGTGG	GGGAGACGTA	GGCGCTTGGG	5880
	GGCCATCTGA	GCGCGGCGGC	GTACCCAAAA	CGCAATACCG	CCGATGACCA	GCACCGCCAG	5940
	GGTACTGCCG	GCCAGCGCGC	CGATGATCAG	GCCCGGGTGG	CTGGGGGCGG	CGGGGGCGTG	6000
	GTGCGGCGCG	ACGTCCTGGA	TCGACGGGAT	GTGCCAGTTT	GGGGGGATCT	GCGAAGACAC	6060
	CGTCCCGGCG	GGATCCTCTA	AGAGGGCCGA	GTCTCGGGG	TCTTCGGAA	CGAGTTCGGG	6120
15	TTGCGTGGCG	TTGGTGGTGT	CGGACAGCTC	CGGCGGCAGC	AGGGTGCTGG	TGTACGGGGG	6180
	CTTGGGGCCG	TGCCACCCGG	CGATTTTAA	GCTGTATAGG	GCGACGGTGC	GCTGGTTTTC	6240
	GGGGATAAAG	CGGGGGAGCA	TCCCAGTGCT	GTGACCGTGC	ACGCCCTGTT	GGTAGGCCCTT	6300
	CGAGGTGAGG	CACGCTGCCG	GGGGGATGCG	CAGGGGGAGA	GCGTACTTGC	AGGAGGCGCG	6360
	GGCCCGGTGC	TCCAAAATAA	ATTGTGTGAT	CTCCGTCCAG	TCGTTTATCT	TCACTAGCCG	6420
20	CAGGTACGTA	CCCGCGGTCT	CGAAGGCGGG	GGCGTGCATC	AGGAATCCCA	GGTTATCCTC	6480
	GCTGACGGCG	CTAAAGCTGT	CATAGTAGCT	CCAGCGGGGC	TGCGTTCGGA	TGGGGCAGAC	6540
	CCCCAACGAC	TTGTTGTAGG	GGCACTCGGT	GTATTCCATA	ACCGTGATGG	GGATAGCGCA	6600
	ATTGTCTCCC	ATGCGATACC	AGGCGATGGT	CAGGTTGTAC	GTGTGCTTTC	GGGCCCTCGTC	6660
	CGAAGCCCCG	CGCACGATCT	GGGGGGCCTC	CGATGGGGCA	TGTAGGAGCA	CGCTGCGGCA	6720
25	GGCACGTTCC	AGCACTGCGT	AGTACACAGT	GATCGGGATG	CTGGGGGGCT	GGAACGGGTC	6780
	CTCCAGGCTC	GGCTGAATGT	GGTAAACACG	CTTCAACCCG	GGGGGGTCGG	TCAGCTGGTC	6840
	CAAAACCGGA	AGGTCTTCC	CGCGAAATCG	ATTGGGATCG	GCCATCTTAA	GCGAGGGGTC	6900
	TGCTAAGGCG	TATTTGGCGC	AGACGACGCG	GAGTCCCACC	GCGACAATA	GCAGGGCCGC	6960
	CGTCCCGACG	CCGGAGGTCA	AACGCCCCAT	GCCGTGATAC	GCGATGCACA	CGAAAAACGG	7020
30	CGACTAGTCG	TTCGCAATGC	AGCTTATGAC	CGAACACCAC	ACCGACCCCG	GGTTTTAAAC	7080
	ACAAAGACTC	TATTATACTC	CTCCTCCTCG	TAAAAATGGA	ACCTCCCCCTC	TGGGGGTGAT	7140
	TTGGTTGCAT	ATGTGGTCGA	ATCGGAGTAT	GGTGGTGCGG	TGGGTCCGCC	ATAACCCCCC	7200
	TTGTGGGATG	GGTTGTGGGC	TTGATGTGTT	TTTAGTTTCT	TTCCCCCCCC	CCCCAAGTTG	7260
	GTACGCTTGG	GGATCGCCGG	ATTATTTTCG	CTTTCGCCGA	CAACCCATGC	CCGGCACGGG	7320
35	GCGTGTGGCA	ACAACCAAAG	TTATATTACC	GACCGCTCCA	TAGCTGCTGT	ACCCCGGGCA	7380
	CCCACACAC	AATCGGGGCG	ATGGGGTGGG	GGCAAGGCCA	GAAAGGCGAA	AAATCATGGG	7440
	GCAAATTGGC	CCGCGTGGGG	GCAGCACCTC	GCCAACTCGC	GACCCAGGCG	ACGCAGGACC	7500
	TCGAGTAGAC	ACGCCATCCC	CAGAATCATG	AGACACAGCC	CCCCCACGAT	GAGGGGGACG	7560
	GCAAAGCCCC	CGGAACGGAT	GAGTGGGGGG	TGCGTGGGGA	GGCGTGCGGT	CGCGTTTGTT	7620
40	GTATCGGACG	CGGGGCCGGT	GGGTGCGGCC	CCAACAGCAG	CACACCCGAG	GATTTCCACA	7680
	ATCCCCCAGG	TCCGAACGGC	ATACCGATCC	ATTGAGACCA	AAACAACAGG	CACGCCCCCC	7740
	GGCGGCGGTC	AAGGTTTTTG	TTTTTGTTTCG	GGGACCCGGG	TGACTTCGTC	TGGGGGCCCTT	7800
	TCTCTGTGTG	GCCAAAAGTT	GCGCGTCTCG	AGGGCCCCCG	GACACGTCTT	TTAAAAGACT	7860

CTGTCACCCA CTGATACCTC CCACCCCCCG CCCCCAGTCC CACAAAACAC AACCAAACCTC 7920  
 ACATATCCGA TTGACGTCAC AGGTTTATTG TTCTTATCGT GGCATTGCGT CGCTGTTCCC 7980  
 TTTCCCGTCC TTCATCGTTT CTCGCCCCCA CCCCACCCCT TAATCCCGCT CGGGTGGCAG 8040  
 ACATACGTAA CGCACGCTCG GGTGCGCGTA TCGCCTGCGC CCCGCCCGGC CGCGCCAAAG 8100  
 5 TTGTGCTGCC AAGGCGACCA GACAAACGAA CGCCGCCGTG TGGATGGTGG TGCTGATGAT 8160  
 AAAGAGGATA TCTAGAGCAG GGGAGGCCGT TAGGAACCAG AACAGGGGGA TGTGTTGGGG 8220  
 TGTGGGGCCC GAGGGCATGT CCTTAGCGGG AGCTTGGGCG GGGGGGCGAG GCGTGTGGG 8280  
 GCGGAGCGGC CCAAGAATTC CTGGCGGGAG CGTGGGGCGG ATGGGCCCGG GCGCGCGGG 8340  
 GGGTGGTTTG TTGGGGTTCG GAGTTCGGAA GCGGAGGCCG GTGGCGCTGT TGTTGTCATC 8400  
 10 GGGGGGTTTC CCGTCCCCCG CGCCCTCAA A CTCCTCGGGT CCGCCCGCAT GTTCGGGGGG 8460  
 TGGGGGGGCT GCGGAGCCCG GGGGAGCGTC CGCGGGTCCG TGTGGGTGCG TCTTTGGGTC 8520  
 CGTTGGGGGG GTACGGGCGG TGCCCGGGT TCCGGGCGTG GCGGTGGTCC CGGCAACCGA 8580  
 AACGTTGGCG GCCGAGGGCC CCGGCGCGGT ACCGGGGGGC GAAGCGGTGA GGGGGGAATC 8640  
 GGCCGTGGGT GCGGCGGAAG CGCCACCGG ACCCGGGGTT GCGGTGCCG GAGGGGTGT 8700  
 15 TTGGGGCCCC GGATTCCTGG GCGGGGGGT CACGTGGGTA AACGTGGGCG GGGGGGTCTG 8760  
 GGGGGCTGGT GTGGTGGGGG GCGTTTCCGC TCGGGGGGCG CTGCTGGTGT TCGTGTGCCC 8820  
 GGCCCCGGGC GTTGCCGCCG CGGCGGGGAT TGGCTACTAC TCCACGGATG CATTCCTGGG 8880  
 CGGGGATGCA ACTGCCGTTT CCTCCGGCGT AACGGCGACC GTTGCGGCTT GTGTGGCCCT 8940  
 CTCGTCGGGG GGAGTATTGG TTGCGGGGGC GGTGCGTCCC CCCCTTGGGT TGACTGATGG 9000  
 20 CCCCATGGCG GTGGGGTAAA GAGGGAGGGG GGTTTTTTGG AGAGGGGAAG TTGGGGAAGG 9060  
 GGAGGAAGGT TTNTGGGGGA GGGGTAAGAG GGGGGNNNG GNGAGGGGG GGNNAAGGG 9120  
 TGGGGGGGAG GNGGGGGGG GCTGTNCCCA CGCTCCCCC CCCCCCGC CCCGGTTCGC 9180  
 AAGCGGGCTT TCGTACCTAC ACCCAGGGCC CCCGCCT 9218

25 (2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 296 amino acids  
 (B) TYPE: amino acid  
 30 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Ile Arg Arg Arg Gly Asn Val Glu Ile Arg Val Tyr Tyr Glu Ser  
 1 5 10 15  
 Val Arg Pro Ser Arg Ser Arg Ser His Leu Lys Pro Ser Asp His Gln  
 20 25 30  
 40 Glu Phe Pro Gly His His Val Ser Pro Gly Ser Pro Gly Phe Pro Glu  
 35 40 45  
 Ser Pro Gly Asn Arg Glu Phe His Asp Leu Pro Glu Asn Pro Gly Ser

	50		55		60											
	Arg	Ala	Tyr	Pro	Gly	Thr	Arg	Asp	Pro	His	Asp	Pro	His	Gly	Cys	Pro
	65				70					75					80	
5	Gly	Ser	Leu	Asp	Pro	His	Gly	Asn	Pro	Ala	Gln	Pro	Ala	Gly	Leu	Pro
				85					90					95		
	Ser	Pro	Val	Pro	Tyr	Ala	Pro	Leu	Gly	Ser	Pro	Asp	Pro	Ser	Ser	Pro
				100					105					110		
	Arg	Gln	Arg	Thr	Tyr	Val	Leu	Pro	Arg	Val	Gly	Ile	Arg	Asn	Ala	Pro
				115					120					125		
10	Ala	Ser	Asp	Thr	Arg	Ala	Pro	Lys	Arg	Ala	His	Ser	Arg	His	Arg	Ala
				130					135					140		
	Asp	Arg	Pro	Pro	Glu	Ser	Pro	Gly	Ser	Glu	Leu	Tyr	Pro	Leu	Asn	Ala
	145				150					155				160		
	Gln	Ala	His	Leu	Gln	Met	Leu	Pro	Ala	Asp	His	Arg	Ala	Phe	Phe	Arg
15				165						170				175		
	Thr	Val	Ile	Glu	Val	Ser	Arg	Leu	Cys	Ala	Leu	Asn	Thr	His	Asp	Pro
				180					185					190		
	Pro	Pro	Pro	Leu	Ala	Gly	Ala	Arg	Val	Gly	Gln	Glu	Ala	Gln	Leu	Val
				195					200					205		
20	His	Thr	Gln	Trp	Leu	Arg	Ala	Asn	Arg	Glu	Ser	Ser	Pro	Leu	Trp	Pro
				210					215					220		
	Trp	Arg	Thr	Ala	Ala	Met	Asn	Phe	Ile	Ala	Ala	Ala	Ala	Pro	Cys	Val
	225				230					235				240		
	Gln	Thr	His	Met	His	Asp	Leu	Leu	Met	Ala	Cys	Ala	Phe	Trp	Cys	Cys
25				245						250				255		
	Leu	Ala	His	Ala	Ser	Thr	Cys	Ser	Tyr	Ala	Gly	Ser	Ala	His	Cys	Gln
				260					265					270		
	His	Leu	Phe	Arg	Ala	Phe	Gly	Cys	Gly	Pro	Pro	Val	Leu	Thr	Thr	Ser
				275					280					285		
30	Arg	Gly	Gln	Gly	Gly	Trp	Cys	Asn								
				290				295								

## (2) INFORMATION FOR SEQ ID NO:10:

## 35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 85 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Thr Ser Arg Pro Ala Asp Gln Asp Ser Val Arg Ser Ser Ala Ser  
 1 5 10 15  
 Val Pro Leu Tyr Pro Ala Asp Val Pro Ala Glu Ala Tyr Tyr Ser Glu  
 5 20 25 30  
 Ser Glu Asp Glu Ala Ala Asn Asp Phe Leu Val Arg Met Gly Arg Gln  
 35 40 45  
 Gln Ser Val Leu Arg Arg Arg Arg Arg Thr Arg Cys Val Gly Leu  
 50 55 60  
 10 Val Ile Ala Cys Leu Val Val Leu Ser Gly Gly Phe Gly Ala Leu Leu  
 65 70 75 80  
 Val Trp Leu Leu Arg  
 85

15 (2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 amino acids

(B) TYPE: amino acid

20 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Val His Ala Val Asp Ala Pro Ser Gln Phe Val Thr Trp Leu Ala Val  
 1 5 10 15  
 Arg Trp Leu Arg Gly Ala Val Gly Leu Gly Ala Val Leu Cys Gly Ile  
 30 20 25 30  
 Ala Phe Tyr Val Thr Ser Ile Arg Ala  
 35 40

(2) INFORMATION FOR SEQ ID NO:12:

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 337 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

40 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

	Val	Ala	Pro	Pro	Arg	His	His	Arg	Val	Ile	Pro	Glu	Val	Ser	His	Val
	1				5					10					15	
5	Arg	Gly	Val	Thr	Val	His	Met	Pro	Glu	Ala	Ile	Met	Phe	Ala	Pro	Gly
				20					25					30		
	Glu	Thr	Phe	Glu	Thr	Lys	Val	Ser	Ile	His	Ala	Val	Ala	His	Asp	Asp
				35				40					45			
	Gly	Pro	Tyr	Ala	Met	Asp	Val	Val	Trp	Met	Arg	Phe	Asp	Val	Pro	Ser
10		50				55					60					
	Ser	Cys	Ala	Glu	Met	Arg	Ile	Tyr	Glu	Ala	Cys	Leu	Tyr	His	Pro	Gln
	65					70				75					80	
	Leu	Pro	Glu	Cys	Leu	Ser	Pro	Ala	Asp	Ala	Pro	Cys	Ala	Val	Ser	Ser
					85				90					95		
15	Trp	Ala	Tyr	Arg	Leu	Ala	Val	Arg	Ser	Tyr	Ala	Gly	Cys	Ser	Arg	Thr
				100				105					110			
	Thr	Pro	Pro	Pro	Arg	Cys	Phe	Ala	Glu	Ala	Arg	Met	Glu	Pro	Val	Pro
				115				120					125			
	Gly	Leu	Ala	Trp	Leu	Ala	Ser	Thr	Val	Asn	Leu	Glu	Phe	Gln	His	Asp
20		130				135						140				
	Gln	His	Ala	Gly	Leu	Cys	Val	Val	Tyr	Val	Asp	Asp	His	Ile	His	Ala
	145					150				155				160		
	Trp	Gly	His	Met	Thr	Ile	Ser	Thr	Ala	Ala	Gln	Tyr	Arg	Asn	Ala	Val
					165				170					175		
25	Val	Glu	Gln	His	Leu	Pro	Gln	Arg	Gln	Pro	Glu	Pro	Val	Glu	Pro	Trp
				180				185					190			
	His	Val	Arg	Ala	Pro	Pro	Pro	Ala	Pro	Ser	Arg	Pro	Leu	Arg	Leu	Gly
			195					200					205			
	Ala	Val	Leu	Gly	Ala	Ala	Leu	Leu	Leu	Ala	Ala	Leu	Gly	Leu	Ser	Ala
30		210				215						220				
	Trp	Ala	Cys	Met	Thr	Cys	Trp	Arg	Arg	Arg	Ser	Trp	Arg	Ala	Val	Lys
	225					230					235				240	
	Ser	Arg	Ala	Ser	Ala	Thr	Gly	Pro	Thr	Tyr	Ile	Arg	Val	Ala	Asp	Ser
					245				250					255		
35	Glu	Leu	Tyr	Ala	Asp	Trp	Ser	Ser	Asp	Ser	Glu	Gly	Glu	Arg	Asp	Gly
				260					265					270		
	Ser	Leu	Trp	Gln	Asp	Pro	Pro	Glu	Arg	Pro	Asp	Ser	Pro	Ser	Thr	Asn
				275				280					285			
	Gly	Ser	Gly	Phe	Glu	Ile	Leu	Ser	Pro	Thr	Ala	Pro	Ser	Val	Tyr	Pro
40		290				295						300				
	His	Ser	Glu	Gly	Arg	Lys	Ser	Arg	Arg	Pro	Leu	Thr	Thr	Phe	Gly	Ser
	305					310					315				320	
	Gly	Ser	Pro	Gly	Arg	Arg	His	Ser	Gln	Ala	Ser	Tyr	Ser	Ser	Val	Leu

325 330 335  
Trp

5 (2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 226 amino acids  
(B) TYPE: amino acid  
10 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met Arg Ala Gly Leu Val Phe Phe Val Gly Val Trp Val Val Ser Cys  
1 5 10 15  
Leu Ala Ala Ala Pro Arg Thr Ser Trp Lys Arg Val Thr Ser Gly Glu  
20 20 25 30  
Asp Val Val Leu Leu Pro Ala Pro Ala Gly Pro Glu Glu Arg Thr Arg  
35 40 45  
Ala His Lys Leu Leu Trp Ala Ala Glu Pro Leu Asp Ala Cys Gly Pro  
50 55 60  
25 Leu Arg Pro Ser Trp Val Trp Pro Pro Arg Arg Val Leu Glu Thr Val  
65 70 75 80  
Val Asp Ala Ala Cys Met Arg Ala Pro Glu Pro Leu Ala Ile Ala Tyr  
85 90 95  
Ser Pro Pro Phe Pro Ala Gly Asp Glu Gly Ser Glu Leu Ala Trp Arg  
30 100 105 110  
Asp Arg Val Ala Val Val Asn Glu Ser Leu Val Ile Tyr Gly Ala Leu  
115 120 125  
Glu Thr Asp Ser Gly Thr Leu Ser Val Val Gly Leu Ser Asp Glu Ala  
130 135 140  
35 Arg Gln Val Ala Ser Val Val Leu Val Val Glu Pro Ala Pro Val Pro  
145 150 155 160  
Thr Pro Thr Pro Asp Asp Tyr Asp Glu Glu Asp Asp Ala Gly Val Ser  
165 170 175  
Thr Pro Val Ser Val Pro Pro Pro Thr Pro Pro Arg Trp Ser Pro Arg  
40 180 185 190  
Gly Pro Pro Glu Ala Pro Ser Cys Tyr Pro Arg Gly Val Pro Arg Arg  
195 200 205  
Asn Gly Pro Tyr Gly Asp Pro Gly Gly His Tyr Val Cys Pro Arg Gly  
183

210  
Asp Val  
225

215

220

5 (2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 429 amino acids

(B) TYPE: amino acid

10 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Val Tyr Leu Trp Ala Arg Val Gly Gly Trp Leu Gly Tyr Leu Gly Gly  
1 5 10 15  
Thr Trp Thr Pro His Lys Gly Ser Leu Glu Gly Gly Lys Leu Gly Gln  
20 20 25 30  
Phe Ile Gly Arg Glu Arg Gly Ala Arg Thr Ala Val Pro Thr Ile Ser  
35 40 45  
His Arg Ala His Ser His Leu Asp Pro Ser Asp Pro Gly Met Pro Gly  
50 55 60  
25 Arg Ser Leu Gln Gly Leu Ala Ile Leu Gly Leu Trp Val Cys Ala Thr  
65 70 75 80  
Gly Leu Val Val Arg Gly Pro Thr Val Ser Leu Val Ser Asp Ser Leu  
85 90 95  
Val Asp Ala Gly Ala Val Gly Pro Gln Gly Phe Val Glu Glu Asp Leu  
30 100 105 110  
Arg Val Phe Gly Glu Leu His Phe Val Gly Ala Gln Val Pro His Thr  
115 120 125  
Asn Tyr Tyr Asp Gly Ile Ile Glu Leu Phe His Tyr Pro Leu Gly Asn  
130 135 140  
35 His Cys Pro Arg Val Val His Val Val Thr Leu Thr Ala Cys Pro Arg  
145 150 155 160  
Arg Pro Ala Val Ala Phe Thr Leu Cys Arg Ser Thr His His Ala His  
165 170 175  
Ser Pro Ala Tyr Pro Thr Leu Glu Leu Gly Leu Ala Arg Gln Pro Leu  
40 180 185 190  
Leu Arg Val Arg Thr Ala Thr Arg Asp Tyr Ala Gly Val Leu Arg Val  
195 200 205  
Trp Val Gly Ser Ala Thr Asn Ala Ser Leu Phe Val Leu Gly Val Ser

210                      215                      220  
 Ala Asn Gly Thr Phe Val Tyr Asn Gly Ser Asp Tyr Gly Ser Cys Asp  
 225                      230                      235                      240  
 Pro Ala Gln Leu Pro Phe Ser Ala Pro Arg Leu Gly Pro Ser Ser Val  
 5                      245                      250                      255  
 Tyr Thr Pro Gly Ala Ser Arg Pro Thr Pro Pro Arg Thr Thr Thr Ser  
                     260                      265                      270  
 Pro Ser Ser Pro Arg Asp Pro Thr Pro Ala Pro Gly Asp Thr Gly Thr  
                     275                      280                      285  
 10    Pro Ala Pro Ala Ser Gly Glu Arg Ala Pro Pro Asn Ser Thr Arg Ser  
                     290                      295                      300  
 Ala Ser Glu Ser Arg His Arg Leu Thr Val Ala Gln Val Ile Gln Ile  
 305                      310                      315                      320  
 Ala Ile Pro Ala Ser Ile Ile Ala Phe Val Phe Leu Gly Ser Cys Ile  
 15                      325                      330                      335  
 Cys Phe Ile His Arg Cys Gln Arg Arg Tyr Arg Arg Pro Arg Gly Gln  
                     340                      345                      350  
 Ile Tyr Asn Pro Gly Gly Val Ser Cys Ala Val Asn Glu Ala Ala Met  
                     355                      360                      365  
 20    Ala Arg Leu Gly Ala Glu Leu Arg Ser His Pro Asn Thr Pro Pro Lys  
                     370                      375                      380  
 Pro Arg Arg Arg Ser Ser Ser Ser Thr Thr Met Pro Ser Leu Thr Ser  
 385                      390                      395                      400  
 Ile Ala Glu Glu Ser Glu Pro Gly Pro Val Val Leu Leu Ser Val Ser  
 25                      405                      410                      415  
 Pro Arg Pro Arg Ser Gly Pro Thr Ala Pro Gln Glu Val  
                     420                      425

## (2) INFORMATION FOR SEQ ID NO:15:

30

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 392 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

35

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

40

Val Cys Ile Ala Tyr His Gly Met Gly Arg Leu Thr Ser Gly Val Gly

1

5

10

15

Thr Ala Ala Leu Leu Val Val Ala Val Gly Leu Arg Val Val Cys Ala

185



		20		25		30	
	Lys	Tyr	Ala	Asp	Pro	Ser	Leu
			35			40	
	Gly	Lys	Asn	Leu	Pro	Val	Leu
5		50			55		60
	Lys	Arg	Val	Tyr	His	Ile	Gln
	65				70		75
	Pro	Ser	Ile	Pro	Ile	Thr	Val
			85			90	
10	Arg	Ser	Val	Leu	Leu	His	Ala
			100			105	
	Gly	Ala	Ser	Asp	Glu	Ala	Arg
		115			120		125
	Trp	Tyr	Arg	Met	Gly	Asp	Asn
15		130			135		140
	Tyr	Thr	Glu	Cys	Pro	Tyr	Asn
	145				150		155
	Thr	Gln	Pro	Arg	Trp	Ser	Tyr
			165			170	
20	Asp	Asn	Leu	Gly	Phe	Leu	Met
		180			185		190
	Thr	Tyr	Leu	Arg	Leu	Val	Lys
		195			200		205
	Phe	Ile	His	Arg	Ala	Arg	Ala
25		210			215		220
	Ile	Pro	Pro	Ala	Ala	Cys	Leu
	225				230		235
	Thr	Val	Asp	Ser	Ile	Gly	Met
			245			250	
30	Arg	Thr	Val	Ala	Lys	Leu	Lys
		260			265		270
	Pro	Tyr	Thr	Ser	Thr	Leu	Leu
		275			280		285
	Ala	Thr	Gln	Pro	Glu	Leu	Val
35		290			295		300
	Leu	Glu	Asp	Pro	Ala	Gly	Thr
	305				310		315
	His	Ile	Pro	Ser	Ile	Gln	Asp
			325			330	
40	Pro	Ser	Asn	Pro	Gly	Leu	Ile
		340			345		350
	Leu	Val	Ile	Gly	Gly	Ile	Ala
		355			360		365

Ala Pro Lys Arg Leu Arg Leu Pro His Ile Arg Asp Asp Asp Ala Pro  
 370 375 380  
 Pro Ser His Gln Pro Leu Phe Tyr  
 385 390

5

## (2) INFORMATION FOR SEQ ID NO:16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

## (ii) MOLECULE TYPE: peptide

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Val Gly Gly Leu Cys Leu Met Ile Leu Gly Met Ala Cys Leu Leu Glu  
 1 5 10 15  
 Val Leu Arg Arg Leu Gly Arg Glu Leu Ala Arg Cys Cys Pro His Ala  
 20 25 30  
 Gly Gln Phe Ala Pro  
 35

25

## (2) INFORMATION FOR SEQ ID NO:17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12489 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

35

GAAAAGGGGG AAGGTGAGGG ATAGGGAAGG AAGAGGAAGG ATAAGAAGTG AAGAAAAAGG 60  
 GAAGAGAAAA GATAGATATG GGGAGAGGAG GAAGAGAGGG GGTGAAGAAG GGAGAAGAGG 120  
 GAGAGGAGAG GTAAAGAGGG GAGAGGAGGT AGGAGTGGAA GGAAGAAGA GAGGAAAAGG 180  
 GGGGGAGGGA AGAGGGGAGG AGCGGCCGAA GCCGGAATGA CAAACAGACG AAGCGACTGG 240  
 GGGAGATCCC CCCGCCCCG AGGACAGCTT TTCCGGGACC TATCCCCGCC ACCGCCGTAT 300  
 AAGCTCGTCT CCACGGTCGA TATCCCCAC CCCGAGACAC CCCGAGAAC ACCGAGCGGC 360  
 CGACAGGCCA CGGACCCCTA TTGCCGTCGA CACACCACCA GCAATCTCCG CGGATGTGCA 420  
 GCGACNGAC CACACGCCC GAAACATCT GANTTCCCCT ATGACCTTTC CCACCACCT 480

40

	CGCGCGGCGC	AGGCGGCCAG	CGCGAGCGCG	CCCGCAAAGG	TCACCACGGG	AACCCAGATG	540
	TTTTCGGGCC	GTGCGGCCCC	GTCGCCGAG	ACCCGCGGAG	ATCGCTGGTG	GTGCGTCGAC	600
	GGTAACAGGG	GTCCGTGGTC	TGTTGGCCGC	TCGGCGTCCT	CCGGGGCGTA	TCGGGGCGGG	660
	GGAGACTCGA	CCGCGGAGAC	GACGCAAAAG	CGGCGGTGGC	GGGGACAGGC	CTTGGAAGAG	720
5	CTGTCCCCGG	CGGCGGTTGT	CCCAGGCCCC	GCCACCACCG	CCGCCGTCAG	GGCCGCGGCG	780
	GCGCGGGGGT	CAATCCACCA	GTAGCGCGCA	CTGCCGCCCTA	TCGTCAGCCC	GCCGACGACA	840
	ACCCCCAGGG	TGCCGACAAA	CAGGGGCCGC	GGGGCGAGGA	GGAAAAGGGG	ATGGACGTCG	900
	TCGGCCAGCA	CCAGCAGCGC	CGCGTAGGCG	GCCCCGAGCG	CCAGGCACTG	GGTCGCCGGG	960
	CCCGGAACCC	CCGGAGGCGC	GCCGGCCTCC	CCGAGGCTCC	ACAGGGCCAC	GGCCGCTCCC	1020
10	CCGCCGACCA	GCCTCAGCCA	CGCGCACACA	CGCGTCGGCG	GCCGGGCGAA	CGCGGGGGGG	1080
	GCGCGGAGGA	ACCCAGGGCC	GGTGCTCCCG	ATCACGAACG	CCCCGACAT	GGCGTCCGCA	1140
	TACGGGCCGC	GCGACAGAAC	GACAGACCCC	ACGAAGCCCA	GCGTGGTGGC	CTGGAGCCAG	1200
	AGATGCGCGA	ACCCCCGAGC	GATGGCACCC	ACGACGGCGA	GGCGGGGGCG	CCCGTCGTCC	1260
	CCATCATCCC	CTCCGAGGCA	GTGTTCGCCG	CTCCCGCCCG	CGACCCCGGG	GCTGTCCCCG	1320
15	CGACACATCC	CGCACCCGGG	GCCCCGACGG	GCTCTCGGTG	AAGCTGCGAA	GGGCCCCGGG	1380
	CCCGCTCTTA	TAAGCGCACG	CAAAACAAAA	AAGGAGGGGG	AAGGGGGGTG	GAAAGGACGG	1440
	AGGCGCAACC	CGGAACCACA	CACAACAGCC	ATATTGGTTG	GAGGGGGGGG	ACACACTACC	1500
	CCATTTATTT	ATTTTTTTAA	CACAACGCAC	CCCGCGTGCC	CGGGCGCGGT	GAACCGTTTCG	1560
	GCCACCACCC	GTCGCCGTCA	GGGCAACCCA	AAACCGTATG	GGGGTCTTTG	GGGACCCCGG	1620
20	AAGGCGGAGA	GGGGGCTGGG	GCTCGCGTCG	CCGGTCTGGA	GGTCGCGAAA	GTACCACGCG	1680
	TAGCGGCCGC	CGTCCCCACC	ACCCTCCGCC	GCCTCGGGTC	CGTATCTCGC	GGAGAGGGGG	1740
	GGGACTCGG	CTCGCGTGGG	GGCGGCGGGC	ACGCCCGTCT	TCGGGCGCTT	GGTGGCGTCA	1800
	TCCGCGGGGA	CAACTTCGGC	CCCGGCGTGC	CTCCGCTTGG	TTCCCGGCGG	TTCCGGGAAC	1860
	GCGGGCGGGA	CCCGGGCCTG	GCCCTCGAGG	CCTTCTTCTT	CTGGGGCGCC	GCGGTGCGCC	1920
25	GCGTCCGGCT	CGGAGGAGAA	GTCCTGGCTG	TCGGTGGGGC	CGGCCCAGAA	GTCTTGGGAC	1980
	GCCAAGCTCC	CGGTGAGGGA	ACCCGGGGTC	CGGCCGAGCC	AGTTCAGGCA	GACCCGCTGG	2040
	GGTTTTAAAA	GAAACACCGC	CGACACCGCG	TTGGGGCCGG	TGGCGGTGAC	ACACACGCTG	2100
	GGGACGTCGG	CCGTGAGGAA	GAAATTGAGG	GTCCCCCGCG	CGACCTGGAG	CCGCCGGAGG	2160
	ACCGCCCGCA	TGCTGCAGTC	GTCGACGACC	ACCGAGAATG	TGCGGTGTGT	GTTTTCCCCG	2220
30	TAGACCGTCT	TGGCGTTGGC	GGCCGCCTGG	CCCGCCTTCT	TCAGCGCGCT	GGTCAGAATC	2280
	TGGACCTGGG	CGCTGGTGCT	GGACGACGCG	CCCTCCTCGC	GGGCGGCAAA	GGTGACGCAG	2340
	GTGCGCGCGT	TAAACACGGA	AAACTTGCCG	TTGGGGCCGA	GCTCGAACGT	GGTGGGTTTG	2400
	GCGGTCTCGT	CCCCGACGGC	GTPCACCACC	TTCGTGAGCT	GGGGCTTCGT	GAGGCGCAGC	2460
	TGGACGTCGG	GGTCGCCCTG	GGGGAGTAGT	ACCGCGAAGC	TCGTCAATC	GCGTTTCATG	2520
35	AGCGTCTCGC	TGGCAAGCTC	CACGGCCTCT	CCGTGCGACG	CGGTGCTCCA	TATGCGCTGC	2580
	ACCAGCGTGC	GAAACGGGGC	CTGGCCCCGT	ACCGTCAGCT	CCACCCGCCG	CAGGTCAGGG	2640
	TACTGGTTGG	CGCGAAACAC	GCTCAGCAGG	GATCGCTTCT	GGTCCACGAG	AGACAGGAAC	2700
	GCCGCGGTGG	GTCCGCCCCA	TCGATAGCGA	CTGAATGCG	AATGGTCGAG	GGGCAGAAAC	2760
	ACCTGCTCGC	CGAAAATCGC	GTTATGTACA	AGGATGCCTC	GGTCGCCCCA	GACCAGGAGC	2820
40	GAGTCCAAAA	GGCTCGTGCG	AAGCGGCGCA	AAGGCCTGCA	GGATCCCGTT	CAGCTCGGCG	2880
	CCCTGCAGCA	CTATCTGGCA	GGGCGGCCAG	TCTTCCGTCC	GCTCGCGCGG	CGACGGGATC	2940
	GCGTCTCCG	AAAGGGGGGC	GGCGGCCGCA	CCGCCGGGAA	GATGAGCCAT	GCCGCGACGC	3000
	TCCCGGGACA	ATCCGCAGAG	ACGAGGCGCG	TCGTGTCACC	GGGCCCCGAG	GCGCGGCCGT	3060

	TTGTGTCGCA	GGCGGAGGGG	GCGGATGACG	CGGACCGGAT	GGGGGTTAGG	GGGGCCGGGG	3120
	GACCCGAGCC	ACAGAGCAGT	GGCTACCCGA	GCCAAGGACT	ACGGCGGACC	CGCCGCCCTA	3180
	GTTTGGTTAA	ATACGCCTTC	CGCTAGTTAG	GCCACACCTT	CTTTGAGGGC	TCGGGGGAGG	3240
	GGGAGGGGGG	GAAGAGAGAG	ATGGTCGGCC	TGCACCGGCG	CGCGCCGGCG	GTTGCACCAA	3300
5	TCCGCACGTA	GATGGGAAAT	AAAAAAGAAT	TATAAAGAGC	GTGCCTTTCC	CGGGATAGCG	3360
	TCTTGTTGGA	GCGGGGTCGT	CGCCGCAGCC	ACTGTACACA	GGGGCGGCGG	GCTTGGGTGT	3420
	CCCGGACCGT	CACACCTATA	CAGCTCTGTA	GAGAGACCTA	TCCGCACCTA	CAATCGTGCC	3480
	GGAATGGGTC	TGTTTGGCAT	GATGAAGTTT	GCCCAGACTC	ACCATCTGGT	GAAGCGCCGG	3540
	GGCCTCCGGG	CCCCGGAGGG	GTACTTTACC	CCCATCGCCG	TGGACCTGTG	GAATGTCATG	3600
10	TATACCCTGG	TGGTTAAATA	TCAGCGCCGC	TACCCAAGTT	ACGACCGCGA	GGCAATCACG	3660
	CTACACTGTC	TCTGTAGTAT	GTTACGGGTG	TTTACCCAAA	AGTCCCTGTT	CCCCATCTTC	3720
	GTGACCGATC	GCGGGGTCGA	GTGTACCGAG	CCGGTTGTGT	TCGGGGCCAA	GGCGATCCTG	3780
	GCCCGCACGA	CGGCCAGTG	CCGCACGGAC	GAGGAGGCCA	GTGACGTAGA	CGCCTCGCCG	3840
	CCGCCTTTCC	CCCATCACCG	ACTCCAGGCC	CAGTTTCCCC	CTTTCCAACA	TGCGCCGGCCG	3900
15	CGGGCACGCC	TTCGCCCCCG	GGGACCGGGG	GAACGCGGGC	CGCCGGCCCA	GGCCCCGGCG	3960
	CCCCCTGGGG	CGCGCCCTCG	AAGCCGGCCC	TGCGCCTGGC	TCACCTGTTC	TGTATCCGCG	4020
	TTCTGCGGGC	GCTGGGGTAC	GCCTACATCA	ACTCGGGTCA	GCTGGAGGCC	GACGACGCCT	4080
	GCGCGAACCT	CTATCATACC	AACACGGTCG	CGTACGTGCA	TACCACGGAT	ACCGATCTCC	4140
	TGCTGATGGG	CTGCGATATC	GTGTTGGACA	TCAGCACCGG	CTACATTCCG	ACGATTCACT	4200
20	GCCCGGACCT	GCTGCAGTAC	TTCAAGATGA	GTTACCCGCA	GTTCTTGCGG	CTGTTCTGTC	4260
	GCTGCCACAC	AGACCTGCAC	CCCAATAACA	CCTACGCGTC	CGTCGAGGAC	GTGCTGCGCG	4320
	AGTGTCACCT	GACCGCCCCG	AGCCGATCCC	AGGCCCGCCG	GGGGGCCCGG	CGGGAGCGCG	4380
	CCAACTCGCG	CTCCCTGGAG	AGCATGCCTA	CGCTGACCGC	GGCCCCGGTC	GGCCTCGAGA	4440
	CGCGCATCTC	GTGGACCGAA	ATTCTGGCCC	AACAGATCGC	GGGCGAGGAC	GACTACGAAG	4500
25	AAGACCCCCC	CCTCCAGCCC	CCGGACGTCG	CCGGTGGGCC	GCGCGACGGC	GCCCCGTCGT	4560
	CCTCCTCGGA	GATACTCACC	CCGCCCAGAC	TCGTGCAGGT	CCCCAACGCG	CAGCGGGTCG	4620
	CGGAACACCG	CGGCTATGTC	GCCGGACGTC	GCCGCCACGT	CATCCACGAC	GCCCCGGAGG	4680
	CCCTGGACTG	GCTGCCCGAT	CCGATGACCA	TCGCCGAGCT	GGTGGAGCAC	AGATACGTCA	4740
	AGTACGTCAT	ATCGCTTATC	AGCCCCAAGG	AGCGGGGACC	CTGGACTCTT	CTAAAAAGAC	4800
30	TGCCCATCTA	TCAGGACCTC	CGCGACGAAG	ATTTAGCGCG	CTCCATCGTG	ACTCGGCATA	4860
	TCACCGCCCC	GGACATCGCC	GACCGGTTTC	TGGCGCAGCT	GTGGGCCCAC	GCGCCCCCGC	4920
	CCGCGTTTTA	CAAGGACGTC	CTGGCTAAAT	TCTGGGACGA	GTAGCCGGAA	CGGAGGAAAC	4980
	GCGCGCCCCC	ATCCCCTCCC	GATGCCCGAC	CTGTTAATAA	TAAGAGTAAT	AAAATCGTTT	5040
	GTTATTATGC	ATCTCGGGGT	TCTGGTCGGC	GCTTGATTTA	TCGGTTGGAC	GCGTTTCCCT	5100
35	TTTGGTCCTT	TTCTCTGGTT	TCGGGCGTTC	CTTCCCTTTC	CCCAGCCGCC	ACCCCCCTCC	5160
	CCTGCGTAAT	AATCACACCG	GAGACCCAAC	AGTCCGTTTC	GACCCCTTTA	TTTCGGTTAG	5220
	ACATCGCTAC	AAGGGCGCCC	AGACCCTCAC	AGATCGTTGA	CGACGGCCCC	GGCGTACGAG	5280
	GTGCTGCGGC	ACTCGAAGAA	GTTGGTGTGT	TTGTGCGTGG	ACATGAGGCT	GAGGGGAAAG	5340
	CTGGCGTCGG	GGGCGGGGGC	GGAATACAGG	GGCTGCATAT	GGATCAGGCC	CAGCAGGCCA	5400
40	TCCGCGCTGA	ATCGCACGTA	GTTCTCGATG	GCCGCCAGGG	CCCCCGGACT	CAGGATAGAG	5460
	CTGTCCGTCG	GGGCCTGGGA	TCGGATGAAC	CCGATCTCGA	TATCCACCGC	CTCCCGAAAC	5520
	AGCCGGTACA	CGCGCGCCGC	CTCGGGCTTG	GCGTGGCCCC	CGAGGTAGTT	GTTGTAGATG	5580
	TAGCACGAGG	CTGTCGTATG	CACGGCCTCG	TCGCGGCTGA	TGAGGTCGTT	CGACTGGCAG	5640

	GTGACCCGCA	GGAGGTTGTT	GGTGCGCAGG	TACGCGATGG	CGGCGAACGA	GGCGGCAAAA	5700
	AAGACGCCCT	CGATGAGGAT	CATGAGGATG	AAC'TTCTCCG	GGATCGAGTC	GCATTCCCGC	5760
	ACCCGCGCCT	CCAGCCAGTC	CACCTTGACG	CGAATGGCCG	GGTGGTTGAT	GGTGCGGGCC	5820
	ACATAGGCGC	GGCGCGCCTG	GTCGTTGTTG	TGAAAGAGCA	CCAGCTGGAT	GATGTTGTAG	5880
5	ACGCGCGAGT	GGACGACCTC	GATGCATTCC	TGCTCCACGT	AGTAGTGAAG	AATGTCCTTC	5940
	TGTTCAAGA	GGCCGAGAG	GCCGCCAGG	TTTTCCGTCA	CCAGGTCGTC	CGCGGCCGAC	6000
	AGGAAGGCAA	ACAGAAAGCG	GTAGAAGCCG	AGCTCGCCCT	CGGAGAGCTT	GGAGACGTCC	6060
	TCCTCGTCCC	CCACGAACAC	GAGCTCGGTC	TCCAGCCAGC	GGTTCAGGAT	GCTGAGGGAG	6120
	CGAAGGTGGT	TGATGTCGGG	GCACTGGGAG	GTGTAGAAGT	ACCGCTCGGG	GGTGGGGCAC	6180
10	ACCGGAATCG	GGGCCGCCCC	GGCCCCGAC	GCGTGGGTAT	CTAGGGGGTC	GGTGCTCGCG	6240
	GGGGAGACGG	CGGGATCCAT	GGCGATATGC	GGGACCGAGA	GCGACGCCTG	ACCCCGATCG	6300
	GAGCGCTGTT	GCTTACAGCG	CGCAGCTTGT	GCAGACGATG	TTGTGTCGTC	CGGCGAACAC	6360
	CCCGCTGTTG	GTCGCCTTGC	GAACCTTGCA	GTAGTACATC	CCCGTCTTCA	GGCCGCGCTT	6420
	ATATGCGTGG	ACGAGAAGGC	GGACCAGGGT	GGAGGCGGGG	AGCGTCCCGT	CCGCCTTCTC	6480
15	TGTGACATAC	AGAGTCATGG	ATTGGCTGTG	ATCAACATAG	GGGGCGCGGT	CTGCACACAG	6540
	GTCGATCAGC	AGTTCCTGGT	CGTAGTCGAA	GGCCGTCTTG	AACCGCCGGA	GGGGGTGGGC	6600
	GGGTCCAGG	CAAGGCAGGG	CCTGGGCCAC	AGACCACTGC	TTGGCCTCGA	GCCCGTCCAT	6660
	CGCGTCCAGG	AGCCGCTTCC	CGCCGAACGT	GCGCTCGAGT	TCCTTCAGCA	AGAGCGTGTT	6720
	GGGGCGCAGC	GTCTCGCCGT	CCCTGGTCAC	CTTGCTGAAC	AGGTTGGTGA	ACAGGGGGGC	6780
20	AAAGCCCTCG	CTGACGTCCG	AGATCTGGGC	CGAGGCGGCG	GTGGGCATGA	GCGCGATGAA	6840
	CTGGCTGTTG	CGCAGGCCGT	GTTTCATCAT	GCTCTGGCGT	AGCATCTCCC	ACTCGCCCTC	6900
	GTACCGCGGG	CTGGCGTTCG	AAAAGCGCTC	CCAGTGAAGG	CGGCCGGCCC	GGTACATGCT	6960
	GCGCTTAAAG	TGGCTGAAGG	GACGCGCCCC	GCGAACGCAC	AGCGCGTTAC	TGGTCTTCAT	7020
	GGCCGCGAGC	AGCATCACCT	CGGCGATGTG	TGTGTTTCAGG	TCCCGGAACT	CGGCCGACTC	7080
25	CAGATCCAGG	CCCATCTTCA	GGCACGCCGT	GTGCAGGCCC	TGCATGCCAA	TGCCCATGGA	7140
	CCGAGGTTG	TCGTGGCCGC	GGGCGCACTG	GGGCGTCGGC	TGCAGCGTGC	TGTCTATCAT	7200
	GATATTAACC	ATTAGCACGC	ACGCCGTGCAC	GCGCTCGCGG	AGCATGCCAA	AATCGAACGT	7260
	CCGCCGGGAG	ACGCATCGGG	CCAGATTAC	GCTGCCAGG	TTGCAGACCC	CGCTGGAGCG	7320
	TTTGGAGGAC	GGGTGGACGA	TTTCCGTGCA	GAGGTTGGAG	CCGGCAATGG	CCGCCCTTG	7380
30	CGTGTTGTAG	ATGTAGTGGC	GGTTTACCGC	GTCCTTAAAC	ATGATGAAGG	GGCTTCCGGT	7440
	GGTGGCCGCG	CTGCGCACGA	TGGCGTACGC	CAGGTCTCGG	ATGGGGATCG	TTTCGCCGAA	7500
	CCCCATGGCC	TCGAGGTGCT	CGTACAGCTT	CTCGAACTCC	TCGCCGTGAA	AGTCGGCGAG	7560
	CGACATGCTG	GTGTCCCGGT	CGAACAGGGA	CCAGGTGACG	TTTTCTCGC	CGTCGAGGTG	7620
	GCGGATCAGG	CGCTTGAAGA	ACAGGTCCGG	CATCCAGAGG	GCGCTGAAGA	TGTTGTCGCA	7680
35	GCGCTGGGCC	TCCTCGCCGG	CGAGGACGCC	CTTCATTCTG	AGCACGGCCC	GAACGTCGCT	7740
	GTGCCAGGGT	TCCAGGTACA	CGCACGCCCC	GGTGGGGCGC	GTGCTCTGTT	TGTTGTGCGC	7800
	CGCCACCAGG	GAGTCCAGGA	CCTTCAGGGC	CGGCATGATG	CTGGCGGTGC	CGGGGCTGGC	7860
	GTCGTTGAAC	GCCTGCATGC	ACAGCCCGAT	GCCCCCGTTG	CGGGCGAGGA	TGGCGCTCAC	7920
	GTGCGCGGTG	ATGGCCCGGA	GGGTGGCCTG	GTTAGTGGTG	GCCTGGGGGT	TTACCAGGTA	7980
40	GCAGCTGGAC	GTGTAGTAGT	TGCGGGTTCC	GAGGTTTCAGC	ATGGCGGGGG	TGGACGGCAC	8040
	GATCTGGTGG	TCGTAGAGGC	GGTGGAAAAA	GAAC'TTGAAC	ATTTCCACC	ACGACCCCTG	8100
	TCGCCCCAGG	GCGATGTGGC	GCATGCCCGG	GGTCGCCCCG	CACGCCAGGA	ACCCGGCGAT	8160
	GCGGGTGTAC	ATCTGGAAGA	CGGACTCCAT	GTAGTGCCCC	CCGAAGCGCT	TGAGGTAAAA	8220

	CTCTTCGTAC	TTCAGCGCCG	ACTGCAGCCC	CCGCTCGACG	AGCGTGTTTCG	GGGTGCTGTG	8280
	GATCAGACAG	TCGTAGGGGT	TCAGGGCCTG	GGCCAGGATC	ATTAGCTGGG	CCTCGTGTTC	8340
	GCGAAGCCTT	TCCGTCAGCC	CGAAGTCCAG	GTCCACCTCC	TTGGAGCGCA	TCCATTCCCTC	8400
	AAAGGAGGCC	TCCCGGGTAC	GGATGCGCAG	GTGAACCAGA	ATCCCCAGGA	TGCGATACAG	8460
5	GCGGGCGGAC	CGCCGCACCA	GGGGTTTGAA	CCCGTTAACC	AGCCGCGTCG	CATACTCCCT	8520
	CAGATGATAG	GGCGTGTATG	CGTTGGGGGG	GACCGGGGGA	AGGTCCAGGC	ACAGGCGTCG	8580
	CATCTCAGCG	AGCGCGTAGT	TCAGGAGCCC	AAAGTCGTCC	TCCGTGAGGC	GGGGGGCGCT	8640
	GCCGAAGGTT	CGTGGGGGCA	CGCGCTTGCT	CTCCTCGCGG	GCGCACCGAC	AGAAGTACTC	8700
	CAGCATGAGC	GCGGGCTCGC	GGTCGACGGC	GTCCCCCAGA	AACCGCGCCA	CCGCTCCGC	8760
10	GTTCTCGGGC	GTGAGTTCTA	GGGGGACTGG	GTAGCCGGGG	TCCGTCCCCA	CAGTGGGCTG	8820
	GCTGTCCAGA	ACCGGCGCCA	CTTCCGCCTG	GGGTGCGGCG	GCGTGGGCGG	CGBAATCGGA	8880
	GTCCGCCGAC	GCGCGCGGGT	CCGTGCGGGA	GCCCCGCCCG	GTGCCGGCGC	CCAGGCCGGG	8940
	GTTTCCGGGG	GACTCCCCGG	GGCGCCGGGG	CTTGGGAAAG	GCCACGGGGG	CAGGGCCGTC	9000
	GCTCCATCTG	CGACGAACGA	CAACGTCGGG	CTGCACGGAG	TCGTCCGACC	GCGAGTCGGA	9060
15	GTGCTGTCA	TCGTGTCAG	TCGCCCCCTG	GGCCCAGATC	GAAGAGGATC	GAGACAGCGT	9120
	CTCCGAACCC	GAGCCCGTAT	CCTCGTCCGA	GGAGTCCGAG	TCCTCCGTTT	CGGAGTCGGA	9180
	CGACGGGCCG	TCTGACCATG	ACTCCGCGGC	CCCAGCTCC	TTCTCGGCGC	CGCCCCCTGGC	9240
	ATCGCGACGG	GCGCAGCACT	CGTGGCCCCA	TGGAAAGGGG	GGAGGAGGGG	GCGGGGGGAC	9300
	AGCCTGGGGT	CCTTGGGGTT	CGGGGGTCCT	TGGGTTCCCG	TGGAGGAAGT	CCCCGGACGT	9360
20	CTGGGTCCCC	ACGGATGTAG	TCGCGGACGG	GCCCCGAGTT	CCGCCGAGCG	CCACGACGGC	9420
	GGTTCGGCCA	TCCCCGCCCG	CTGCGACGTT	TGAGATCGCG	ACGAAGGCGC	CGGTGGACGT	9480
	AGCGCCCTCG	AGGTCACGCA	AATGACCGCG	CGCCACGTCT	CCGTGATTA	TCATACTGCA	9540
	GTTCGAGCCG	CATTGAACAA	AGCTGCTGTC	GCTAATGCGG	TAGGCCGCGG	GGCCGGGGGG	9600
	ATCGCTGGAA	AGCACCATCA	CGCCGCTGAC	TTTCTGCAA	AACACGTGGT	CGCCGCCAGG	9660
25	GGGGGCGACC	TCGGGCTCCC	GGGGTTCTTG	TCGTTGCGAC	GGAGACCGCG	CTCCGGCGAG	9720
	GGCGGATGCG	GCAGGGCGGT	TGGCCATTGG	AACCAAGGTG	ACGGTGGCTC	GGACGCGATG	9780
	AACGGAACAA	GCGGCGGGCC	CGTGGGAGGC	GGGGTGGGCG	GTACGACCGA	AAGGACGCAC	9840
	CGGCCGACGA	ATCACC CGGG	CAGTCGCCAA	CAATCTGTG	ACAGGACAGC	ACCGGCCGAC	9900
	AGGAACGCAA	CAGGAATCGG	TGGGCGAAAA	CGTGAGGGCT	CGCAGGAGAG	TGGCGACCGT	9960
30	CAAGTCCCCG	AAGCACCAC	CTGTTGTGGT	AACCCAAACC	GCCCCGCTTG	GCTTTTATCC	10020
	CCGGCACACA	GGGGTGGGGT	GCGATCGTGT	GCTGAGTCAA	TCTAAGGAGG	ATATGCCTAT	10080
	CTCCTGAGTC	AGGGGGGTGT	AACGTGTCCA	TGAATCCCAT	TTGCATGTG	GTTTCCGCCA	10140
	TTTCAGAGAC	ACACGCCAAA	CACACCACGT	CACGGATTTA	ACCTGGCTTT	ATTGAGACGG	10200
	ACGCACGGGG	CCGCGCACGG	CCAAGACGGC	GAGGGCGGAG	TTGGTGTGTA	GGAGGGGGGG	10260
35	GGGGGAAAGT	GGGAGGGGAA	GGTCGCTCGG	GTTGGGCGGC	CGTCGTCAAT	AGACGATCAC	10320
	GCGCAGGCC	GCCACCCGCC	GGGGCGCCAC	ACCACGCCCT	CCAGGATGAC	AACCGGAACG	10380
	CCGACCGCCC	CAAACCGCTC	GGTGCGGTGC	AGACAGCGGG	GGCTGTTGTC	TGGCATGACA	10440
	CCCAGCTCTG	CGAAGGCGGC	GTATGTGCAG	GTGCGTGGG	TAGGGCGCCC	CCGCAGATCC	10500
	GGCTGCCAGC	GGTACAGCCT	GTCGACAAAC	TGGGGGTTTC	TGACTTGGCT	CGCCGAGCAC	10560
40	CGCGGGGATT	GCGAGCTCGC	GGCCAGGACG	GCGAGCGGAA	AGTCAACGTT	TCGGTTCGGG	10620
	GGAAGGGGCG	TCATGTCCTC	GGCCCCGTGA	ATCGTGTTCG	TTATCCGGAG	GCGCCCGAAC	10680
	AGCCGCTCGA	GGGCCGCCCTC	CAGTCCGCGC	GGGGGAGCT	GGCTCTTGAC	CACGTACACG	10740
	CAACAGAGCT	CCTGGTCCCCG	GCACTGTGCG	TAGGTGAAAA	CCAAGTACAG	GAACGCCGCC	10800

	CCAGGGCCCC	CCAGGCACAG	CTCGGCGTCC	AGGTCCACAA	AGGCGCAGGC	CGGGATGGTA	10860
	ACGGCCGAAC	AGGGCCTGCC	CGGGTGGCCG	GTGTGTGTGG	CCTCCTGGGG	TCCGCTGCAG	10920
	ACGGCGGTGA	CGCTAGCCAG	CTCGTCCTGT	GTGACCACGA	CACTAGCCCC	CCGAAAAACC	10980
	GCATGACCTC	GTGGGGAAC	ACGTGCGTGT	GGACCATGGC	GCGCAGGCAT	TCGGAAAAACC	11040
5	GGTCCAGCCG	CGCCCCGGCC	CGCGTCCCGC	CGTAGTTGGT	CGTGATGTGG	GCCCGGACCG	11100
	CCTCGGCGGC	GTCGCGCGCA	TCGTAGCGCG	CGGCGCACGC	GGCCACCAGA	AAGTTAAACG	11160
	ACACAAGGCC	CGCGCGCGCG	CAGCCGCTCT	CGCCCCGCC	GGACCCAGG	GCGGAGGCCT	11220
	CCAGGAGCTG	CCCCCAGGCC	TCGGCCAGTC	GCTCGGTCTG	CCGGCCGGGC	GGAGCCCGAT	11280
	GGCGGGCCAG	GTGGGGGAGG	TCGGTGGGGT	GCCGAGGGC	CAAAAGGAGC	GCCCCGGCCC	11340
10	GCTCCGCGTT	TGGTTGGCAG	AGATCCGTCA	GGTTACCTG	GCGGGTCAGG	TGATAGCGGG	11400
	GGGCGCCGGG	GACGGTCAGC	CCTCCCGCGC	GCCGGGTCC	CCTGAGGATC	TTGTCCACGG	11460
	CCTGCTCGGT	ATCGTCTCGG	GTAGCGCGCG	TTCCCGGAGA	CGACTCCGCG	GGGTCGATCC	11520
	CCAGCAGCCA	CATCGTACTG	GCCGTCTCC	TGGGGGCCCT	GTCCCCCAA	AGCACCGGGC	11580
	CGTCGCGGCG	AGCGATGGGC	GGGCGCAGGA	CGGCCCGGAG	CGGGGCGTGG	CCCGCGAGCT	11640
15	CGCGCGGGCT	GGTGGTGGTT	TCCACGGCAC	TCTCGGCCCA	CGCCATCGGG	GCTGTCGGGA	11700
	GTGGCTTGGT	TTTCATGGTT	TTCCCGCCGA	TCGCCAACGG	GCGAGTTGTG	GGGGGGGGGG	11760
	TGGTGGAGGT	GGGCCACCCC	CCNGGCCGCC	CGTANCCCTC	CCCCCCCCTT	CTGCCCTCCC	11820
	CCCCCCTCC	TCGTGTCTTC	TACCCATCTC	TCTTCCATT	CCCCTCCCCC	CTCTCCCTCC	11880
	GTCCCTCTGC	CCCTCCCTCC	CTCCTCCCCG	TTCTCTCTCA	TTCCCTCCTT	ATTCTCCAC	11940
20	TATTCGTCTC	TTGGCCTCTC	GCTTATCCCC	ATATCTCCCC	TCCTGTGGTC	CTATTATTCT	12000
	CATCCTATCT	CCTCTCCTTC	TTACCATGAC	AATTCTCTCC	TCATCTCTCC	TCTCCTTACT	12060
	CCCCTCTCTC	TCTCTGTCCA	TGCTCTCCCC	TCGTCTATCT	ACCTCCACAT	CTCCATTCTC	12120
	TCCTCGTACT	CTCTACTCCT	CCCTATCATG	TCATGCGTGC	CCTCGCGTAC	TCTTCTCTCC	12180
	CCTATCTTTC	CACTTCTCCC	CCCCGCTCTA	CTCTTCTCT	CATGCGTCGC	TCTATCTTCT	12240
25	CTCTCCTCCT	TCCGTCTATC	TCTCTCCTCC	CCTCTTTTCT	TCACCCCTCG	GTGACCCCTC	12300
	CCTTGTTTAC	ATCACTTTTG	AAACTCCAGC	TCTTGCTCTC	CCCCACTGCT	CCTCCTCCCC	12360
	TGACCCCCCC	CCCTCTCCCT	TTCTCTCGTA	TCTCAGTCCT	CCTCTCTCCC	CCTCTTTTTT	12420
	CCCTCTCTTT	CGTTCTCCTC	CCTTCTTTTC	GATCCCGCTC	CTCTGTCTCC	TCCTCCTTCT	12480
30	CTTCTCTT						12489

## (2) INFORMATION FOR SEQ ID NO:18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 335 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Val Cys Pro Pro Pro Pro Thr Asn Met Ala Val Val Cys Gly Ser Gly

	1		5		10		15
	Leu	Arg	Leu	Arg	Pro	Phe	His
					Pro	Pro	Ser
					Pro	Ser	Phe
					Phe	Phe	Val
					Leu		
			20		25		30
	Arg	Ala	Leu	Ile	Arg	Ala	Gly
					Pro	Gly	Pro
					Phe	Ala	Asp
					Arg	Ala	Pro
5			35		40		45
	Ser	Gly	Pro	Gly	Cys	Gly	Met
					Cys	Arg	Gly
					Asp	Ser	Pro
					Gly	Val	Ala
			50		55		60
	Gly	Gly	Ser	Gly	Glu	His	Cys
					Leu	Gly	Gly
					Asp	Asp	Gly
					Asp	Gly	Asp
					Asp	Gly	Asp
					Gly		
	65				70		75
							80
10	Arg	Pro	Arg	Leu	Ala	Cys	Val
					Gly	Ala	Ile
					Arg	Phe	Ala
					His	Leu	Trp
			85		90		95
	Leu	Gln	Ala	Thr	Thr	Leu	Gly
					Phe	Val	Gly
					Ser	Val	Val
					Leu	Ser	Arg
			100		105		110
	Gly	Pro	Tyr	Ala	Asp	Ala	Met
					Ser	Gly	Ala
					Phe	Val	Ile
					Gly	Ser	Thr
15			115		120		125
	Gly	Leu	Gly	Phe	Leu	Arg	Ala
					Pro	Pro	Ala
					Phe	Ala	Arg
					Pro	Pro	Thr
			130		135		140
	Arg	Val	Cys	Ala	Trp	Leu	Arg
					Leu	Val	Gly
					Gly	Gly	Gly
					Ala	Ala	Val
					Trp		
	145				150		155
							160
20	Ser	Leu	Gly	Glu	Ala	Gly	Ala
					Pro	Pro	Gly
					Val	Pro	Gly
					Pro	Ala	Thr
			165		170		175
	Gln	Cys	Leu	Ala	Leu	Gly	Ala
					Ala	Tyr	Ala
					Ala	Ala	Leu
					Leu	Val	Leu
					Ala		
			180		185		190
	Asp	Asp	Val	His	Pro	Leu	Phe
					Leu	Leu	Ala
					Pro	Arg	Pro
					Leu	Phe	Val
25			195		200		205
	Gly	Thr	Leu	Gly	Val	Val	Val
					Gly	Gly	Leu
					Thr	Ile	Gly
					Gly	Gly	Ser
					Ala		
			210		215		220
	Arg	Tyr	Trp	Trp	Ile	Asp	Pro
					Arg	Ala	Ala
					Ala	Ala	Ala
					Leu	Thr	Ala
					Ala		
			225		230		235
							240
30	Val	Val	Ala	Gly	Leu	Gly	Thr
					Thr	Ala	Ala
					Gly	Asp	Ser
					Phe	Ser	Lys
			245		250		255
	Ala	Cys	Pro	Arg	His	Arg	Arg
					Phe	Cys	Val
					Val	Val	Ser
					Ala	Val	Glu
					Ser		
			260		265		270
	Pro	Pro	Pro	Arg	Tyr	Ala	Pro
					Glu	Asp	Ala
					Glu	Arg	Pro
					Thr	Asp	His
35			275		280		285
	Gly	Pro	Leu	Leu	Pro	Ser	Thr
					His	His	Gln
					Arg	Ser	Pro
					Arg	Val	Cys
			290		295		300
	Gly	Asp	Gly	Ala	Ala	Arg	Pro
					Glu	Asn	Ile
					Trp	Val	Pro
					Val	Val	Thr
			305		310		315
							320
40	Phe	Ala	Gly	Ala	Leu	Ala	Cys
					Ala	Arg	Trp
					Trp	Glu	Arg
					Ser		
			325		330		335

(2) INFORMATION FOR SEQ ID NO:19:



## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 466 amino acids

(B) TYPE: amino acid

5 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met Ala His Leu Pro Gly Gly Ala Ala Ala Pro Leu Ser Glu Asp  
 1 5 10 15  
 Ala Ile Pro Ser Pro Arg Glu Arg Thr Glu Asp Trp Pro Pro Cys Gln  
 15 20 25 30  
 Ile Val Leu Gln Gly Ala Glu Leu Asn Gly Ile Leu Gln Ala Phe Ala  
 35 40 45  
 Pro Leu Arg Thr Ser Leu Leu Asp Ser Leu Leu Val Val Gly Asp Arg  
 50 55 60  
 20 Gly Ile Leu Val His Asn Ala Ile Phe Gly Glu Gln Val Phe Leu Pro  
 65 70 75 80  
 Leu Asp His Ser Gln Phe Ser Arg Tyr Arg Trp Gly Gly Pro Thr Ala  
 85 90 95  
 Ala Phe Leu Ser Leu Val Asp Gln Lys Arg Ser Leu Leu Ser Val Phe  
 100 105 110  
 25 Arg Ala Asn Gln Tyr Pro Asp Leu Arg Arg Val Glu Leu Thr Val Thr  
 115 120 125  
 Gly Gln Ala Pro Phe Arg Thr Leu Val Gln Arg Ile Trp Thr Thr Ala  
 130 135 140  
 30 Ser Asp Gly Glu Ala Val Glu Leu Ala Ser Glu Thr Leu Met Lys Arg  
 145 150 155 160  
 Glu Leu Thr Ser Phe Ala Val Leu Leu Pro Gln Gly Asp Pro Asp Val  
 165 170 175  
 Gln Leu Arg Leu Thr Lys Pro Gln Leu Thr Lys Val Val Asn Ala Val  
 180 185 190  
 35 Gly Asp Glu Thr Ala Lys Pro Thr Thr Phe Glu Leu Gly Pro Asn Gly  
 195 200 205  
 Lys Phe Ser Val Phe Asn Ala Arg Thr Cys Val Thr Phe Ala Ala Arg  
 210 215 220  
 40 Glu Glu Gly Ala Ser Ser Ser Thr Ser Ala Gln Val Gln Ile Leu Thr  
 225 230 235 240  
 Ser Ala Leu Lys Lys Ala Gly Gln Ala Ala Ala Asn Ala Lys Thr Val  
 245 250 255

Tyr Gly Glu Asn Thr Thr Phe Ser Val Val Val Asp Asp Cys Ser Met  
                   260                  265                  270  
 Arg Ala Val Leu Arg Arg Leu Gln Val Gly Gly Gly Thr Leu Asn Phe  
                   275                  280                  285  
 5 Phe Leu Thr Ala Asp Val Pro Ser Val Cys Val Thr Ala Thr Gly Pro  
                   290                  295                  300  
 Asn Ala Val Ser Ala Val Phe Leu Leu Lys Pro Gln Arg Val Cys Leu  
 305                  310                  315                  320  
 Asn Trp Leu Gly Arg Thr Pro Gly Ser Ser Thr Gly Ser Leu Ala Ser  
 10                  325                  330                  335  
 Gln Asp Ser Arg Ala Gly Pro Thr Asp Ser Gln Asp Phe Ser Ser Glu  
                   340                  345                  350  
 Pro Asp Ala Gly Asp Arg Gly Ala Pro Glu Glu Glu Gly Leu Glu Gly  
                   355                  360                  365  
 15 Gln Ala Arg Val Pro Pro Ala Phe Pro Glu Pro Pro Gly Thr Lys Arg  
                   370                  375                  380  
 Arg His Ala Gly Ala Glu Val Val Pro Ala Asp Asp Ala Thr Lys Arg  
 385                  390                  395                  400  
 Pro Lys Thr Gly Val Pro Ala Ala Pro Thr Arg Ala Glu Ser Pro Pro  
 20                  405                  410                  415  
 Leu Ser Ala Arg Tyr Gly Pro Glu Ala Ala Glu Gly Gly Gly Asp Gly  
                   420                  425                  430  
 Gly Arg Tyr Ala Trp Tyr Phe Arg Asp Leu Gln Thr Gly Asp Asp Ser  
                   435                  440                  445  
 25 Pro Leu Ser Ala Phe Arg Gly Pro Gln Arg Pro Pro Tyr Gly Phe Gly  
                   450                  455                  460  
 Leu Pro  
 465

30 (2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 218 amino acids

(B) TYPE: amino acid

35 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Gly Leu Phe Gly Met Met Lys Phe Ala Gln Thr His His Leu Val  
 1                  5                  10                  15

	Lys	Arg	Gly	Leu	Arg	Ala	Pro	Glu	Gly	Tyr	Phe	Thr	Pro	Ile	Ala	
			20					25					30			
	Val	Asp	Leu	Trp	Asn	Val	Met	Tyr	Thr	Leu	Val	Val	Lys	Tyr	Gln	Arg
		35					40						45			
5	Arg	Tyr	Pro	Ser	Tyr	Asp	Arg	Glu	Ala	Ile	Thr	Leu	His	Cys	Leu	Cys
	50						55					60				
	Ser	Met	Leu	Arg	Val	Phe	Thr	Gln	Lys	Ser	Leu	Phe	Pro	Ile	Phe	Val
	65				70						75					80
	Thr	Asp	Arg	Gly	Val	Glu	Cys	Thr	Glu	Pro	Val	Val	Phe	Gly	Ala	Lys
10				85					90						95	
	Ala	Ile	Leu	Ala	Arg	Thr	Thr	Ala	Gln	Cys	Arg	Thr	Asp	Glu	Glu	Ala
			100					105					110			
	Ser	Asp	Val	Asp	Asp	Pro	Pro	Phe	Pro	His	His	Arg	Leu	Gln	Ala	Gln
		115					120					125				
15	Phe	Pro	Pro	Phe	Gln	His	Ala	Pro	Pro	Arg	Ala	Arg	Leu	Arg	Pro	Gly
	130						135				140					
	Gly	Pro	Gly	Glu	Arg	Gly	Pro	Pro	Ala	Gln	Ala	Arg	Arg	Pro	Pro	Gly
	145					150				155						160
	Ala	Arg	Pro	Arg	Ser	Arg	Pro	Cys	Ala	Trp	Leu	Thr	Cys	Ser	Val	Ser
20				165					170							175
	Ala	Phe	Cys	Gly	Arg	Trp	Gly	Thr	Pro	Thr	Ser	Thr	Arg	Val	Ser	Trp
			180						185				190			
	Arg	Pro	Thr	Thr	Pro	Ala	Arg	Thr	Ser	Ile	Ile	Pro	Thr	Arg	Ser	Arg
		195					200					205				
25	Thr	Cys	Ile	Pro	Arg	Ile	Pro	Ile	Ser	Cys						
	210						215									

(2) INFORMATION FOR SEQ ID NO:21:

30 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## 35

- (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

40 Val His Thr Thr Asp Thr Asp Leu Leu Leu Met Gly Cys Asp Ile Val  
1 5 10 15  
Leu Asp Ile Ser Thr Gly Tyr Ile Pro Thr Ile His Cys Arg Asp Leu  
20 25 30  
196

Leu Gln Tyr Phe Lys Met Ser Tyr Pro Gln Phe Leu Ala Leu Phe Val  
 35 40 45  
 Arg Cys His Thr Asp Leu His Pro Asn Asn Thr Tyr Ala Ser Val Glu  
 50 55 60  
 5 Asp Val Leu Arg Glu Cys His Trp Thr Ala Pro Ser Arg Ser Gln Ala  
 65 70 75 80  
 Arg Arg Gly Ala Arg Arg Glu Arg Ala Asn Ser Arg Ser Leu Glu Ser  
 85 90 95  
 Met Pro Thr Leu Thr Ala Ala Pro Val Gly Leu Glu Thr Arg Ile Ser  
 10 100 105 110  
 Trp Thr Glu Ile Leu Ala Gln Gln Ile Ala Gly Glu Asp Asp Tyr Glu  
 115 120 125  
 Glu Asp Pro Pro Leu Gln Pro Pro Asp Val Ala Gly Gly Pro Arg Asp  
 130 135 140  
 15 Gly Ala Arg Ser Ser Ser Ser Glu Ile Leu Thr Pro Pro Glu Leu Val  
 145 150 155 160  
 Gln Val Pro Asn Ala Gln Arg Val Ala Glu His Arg Gly Tyr Val Ala  
 165 170 175  
 Gly Arg Arg Arg His Val Ile His Asp Ala Pro Glu Ala Leu Asp Trp  
 20 180 185 190  
 Leu Pro Asp Pro Met Thr Ile Ala Glu Leu Val Glu His Arg Tyr Val  
 195 200 205  
 Lys Tyr Val Ile Ser Leu Ile Ser Pro Lys Glu Arg Gly Pro Trp Thr  
 210 215 220  
 25 Leu Leu Lys Arg Leu Pro Ile Tyr Gln Asp Leu Arg Asp Glu Asp Leu  
 225 230 235 240  
 Ala Arg Ser Ile Val Thr Arg His Ile Thr Ala Pro Asp Ile Ala Asp  
 245 250 255  
 Arg Phe Leu Ala Gln Leu Trp Ala His Ala Pro Pro Pro Ala Phe Tyr  
 30 260 265 270  
 Lys Asp Val Leu Ala Lys Phe Trp Asp Glu  
 275 280

## (2) INFORMATION FOR SEQ ID NO:22:

35

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 528 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

40

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

	Met	Arg	Ala	Gly	Leu	Val	Phe	Phe	Val	Gly	Val	Trp	Val	Val	Ser	Cys
	1				5					10					15	
5	Leu	Ala	Ala	Ala	Pro	Arg	Thr	Ser	Trp	Lys	Arg	Val	Thr	Ser	Gly	Glu
				20					25					30		
	Asp	Val	Val	Leu	Leu	Pro	Ala	Pro	Ala	Gly	Pro	Glu	Glu	Arg	Thr	Arg
		35					40					45				
	Ala	His	Lys	Leu	Leu	Trp	Ala	Ala	Glu	Pro	Leu	Asp	Ala	Cys	Gly	Pro
10		50					55					60				
	Leu	Arg	Pro	Ser	Trp	Val	Trp	Pro	Pro	Arg	Arg	Val	Leu	Glu	Thr	Val
	65				70					75					80	
	Val	Asp	Ala	Ala	Cys	Met	Arg	Ala	Pro	Glu	Pro	Leu	Ala	Ile	Ala	Tyr
				85						90				95		
15	Ser	Pro	Pro	Phe	Pro	Ala	Gly	Asp	Glu	Gly	Ser	Glu	Leu	Ala	Trp	Arg
				100				105					110			
	Asp	Arg	Val	Ala	Val	Val	Asn	Glu	Ser	Leu	Val	Ile	Tyr	Gly	Ala	Leu
		115					120					125				
	Glu	Thr	Asp	Ser	Gly	Thr	Leu	Ser	Val	Val	Gly	Leu	Ser	Asp	Glu	Ala
20		130					135					140				
	Arg	Gln	Val	Ala	Ser	Val	Val	Leu	Val	Val	Glu	Pro	Ala	Pro	Val	Pro
	145					150				155					160	
	Thr	Pro	Thr	Pro	Asp	Asp	Tyr	Asp	Glu	Glu	Asp	Asp	Ala	Gly	Val	Ser
				165					170					175		
25	Thr	Pro	Val	Ser	Val	Pro	Pro	Pro	Thr	Pro	Pro	Arg	Gly	Pro	Pro	Val
				180					185				190			
	Ala	Pro	Pro	Thr	His	Pro	Arg	Val	Ile	Pro	Glu	Val	Ser	His	Val	Arg
		195					200					205				
	Gly	Val	Thr	Val	His	Met	Pro	Glu	Ala	Ile	Leu	Phe	Ala	Pro	Gly	Glu
30		210					215					220				
	Thr	Phe	Gly	Thr	Asn	Val	Ser	Ile	His	Ala	Ile	Ala	His	Asp	Asp	Gly
	225				230					235					240	
	Pro	Tyr	Ala	Met	Asp	Val	Val	Trp	Met	Arg	Phe	Asp	Val	Pro	Ser	Ser
				245					250				255			
35	Cys	Ala	Glu	Met	Arg	Ile	Tyr	Glu	Ala	Cys	Leu	Tyr	His	Pro	Gln	Leu
			260					265				270				
	Pro	Glu	Cys	Leu	Ser	Pro	Ala	Asp	Ala	Pro	Cys	Ala	Val	Ser	Ser	Trp
		275					280					285				
	Ala	Tyr	Arg	Leu	Ala	Val	Arg	Ser	Tyr	Ala	Gly	Cys	Ser	Arg	Thr	Thr
40		290					295				300					
	Pro	Pro	Pro	Arg	Cys	Phe	Ala	Glu	Ala	Arg	Met	Glu	Pro	Val	Pro	Gly
	305				310					315				320		
	Leu	Ala	Trp	Leu	Ala	Ser	Thr	Val	Asn	Leu	Glu	Phe	Gln	His	Asp	Gln

```

                    325                    330                    335
His Ala Gly Leu Cys Val Val Tyr Val Asp Asp His Ile His Ala Trp
                    340                    345                    350
Gly His Met Thr Ile Ser Thr Ala Ala Gln Tyr Arg Asn Ala Val Val
5                    355                    360                    365
Glu Gln His Leu Pro Gln Arg Gln Pro Glu Pro Val Glu Pro Trp His
                    370                    375                    380
Val Arg Ala Pro Pro Pro Ala Pro Ser Arg Pro Leu Arg Leu Gly Ala
385                    390                    395                    400
10 Val Leu Gly Ala Ala Leu Leu Leu Ala Ala Leu Gly Leu Ser Ala Trp
                    405                    410                    415
Ala Cys Met Thr Cys Trp Arg Arg Arg Ser Trp Arg Ala Val Lys Ser
                    420                    425                    430
Arg Ala Ser Ala Thr Gly Pro Thr Tyr Ile Arg Val Ala Asp Ser Glu
15                    435                    440                    445
Leu Tyr Ala Asp Trp Ser Ser Asp Ser Glu Gly Glu Arg Asp Gly Ser
                    450                    455                    460
Leu Trp Gln Asp Pro Pro Glu Arg Pro Asp Ser Pro Ser Thr Asn Gly
465                    470                    475                    480
20 Ser Gly Phe Glu Ile Leu Ser Pro Thr Ala Pro Ser Val Tyr Pro His
                    485                    490                    495
Ser Glu Gly Arg Lys Ser Arg Arg Pro Leu Thr Thr Phe Gly Ser Gly
                    500                    505                    510
Ser Pro Gly Arg Arg His Ser Gln Ala Ser Tyr Ser Ser Val Leu Trp
25                    515                    520                    525

```

## (2) INFORMATION FOR SEQ ID NO:23:

```

30 (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 1160 amino acids
    (B) TYPE: amino acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

```

```

35 (ii) MOLECULE TYPE: peptide

```

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```

40 Val Ile Arg Arg Pro Val Arg Pro Phe Gly Arg Thr Ala His Pro Ala
   1                    5                    10                    15
Ser His Gly Pro Ala Ala Val Ser Val His Arg Val Arg Ala Thr Val
   20                    25                    30
Thr Leu Val Pro Met Ala Asn Arg Pro Ala Ala Ser Ala Gly Ala Arg

```

35 40 45  
 Ser Pro Ser Gln Glu Pro Arg Glu Pro Glu Val Ala Pro Pro Gly Gly  
 50 55 60  
 Asp His Val Phe Cys Arg Lys Val Ser Gly Val Met Val Leu Ser Ser  
 5 65 70 75 80  
 Asp Pro Pro Gly Pro Ala Ala Tyr Arg Ile Ser Asp Ser Ser Phe Val  
 85 90 95  
 Gln Cys Gly Ser Asn Cys Ser Met Ile Ile Asp Gly Asp Val Arg His  
 100 105 110  
 10 Leu Arg Asp Leu Glu Gly Ala Thr Ser Thr Gly Ala Phe Val Ala Ile  
 115 120 125  
 Ser Asn Val Ala Ala Gly Gly Asp Gly Arg Thr Ala Val Val Gly Gly  
 130 135 140  
 Thr Ser Gly Pro Ser Ala Thr Thr Ser Val Gly Thr Gln Thr Ser Gly  
 15 145 150 155 160  
 Glu Phe Leu His Gly Asn Pro Arg Thr Pro Glu Pro Gln Gly Pro Gln  
 165 170 175  
 Ala Val Pro Pro Pro Pro Pro Pro Phe Pro Trp Gly His Glu Cys  
 180 185 190  
 20 Cys Ala Arg Arg Asp Arg Gly Ala Glu Lys Asp Val Gly Ala Ala Glu  
 195 200 205  
 Ser Trp Ser Asp Gly Pro Ser Ser Asp Ser Glu Thr Glu Asp Ser Asp  
 210 215 220  
 Ser Ser Asp Glu Asp Thr Gly Ser Gly Ser Glu Thr Leu Ser Arg Ser  
 25 225 230 235 240  
 Ser Ser Ile Trp Ala Ala Gly Ala Thr Asp Asp Asp Asp Ser Asp Ser  
 245 250 255  
 Asp Ser Arg Ser Asp Asp Ser Val Gln Pro Asp Val Val Val Arg Arg  
 260 265 270  
 30 Arg Trp Ser Asp Gly Pro Ala Pro Val Ala Phe Pro Lys Pro Arg Arg  
 275 280 285  
 Pro Gly Asp Ser Pro Gly Asn Pro Gly Leu Gly Ala Gly Thr Gly Pro  
 290 295 300  
 Gly Ser Ala Thr Asp Pro Arg Ala Ser Ala Asp Ser Asp Ser Ala Ala  
 35 305 310 315 320  
 His Ala Ala Ala Pro Gln Ala Glu Val Ala Pro Val Leu Asp Ser Gln  
 325 330 335  
 Pro Thr Val Gly Thr Asp Pro Gly Tyr Pro Val Pro Leu Glu Leu Thr  
 340 345 350  
 40 Pro Glu Asn Ala Glu Ala Val Ala Arg Phe Leu Gly Asp Ala Val Asp  
 355 360 365  
 Arg Glu Pro Ala Leu Met Leu Glu Tyr Phe Cys Arg Cys Ala Arg Glu  
 370 375 380

Glu Ser Lys Arg Val Pro Pro Arg Thr Phe Gly Ser Ala Pro Arg Leu  
 385 390 395 400  
 Thr Glu Asp Asp Phe Gly Leu Leu Asn Tyr Ala Glu Met Arg Arg Leu  
 405 410 415  
 5 Cys Leu Asp Leu Pro Pro Val Pro Pro Asn Ala Tyr Thr Pro Tyr His  
 420 425 430  
 Leu Arg Glu Tyr Ala Thr Arg Leu Val Asn Gly Phe Lys Pro Leu Val  
 435 440 445  
 Arg Arg Ser Ala Arg Leu Tyr Arg Ile Leu Gly Ile Leu Val His Leu  
 10 450 455 460  
 Arg Ile Arg Thr Arg Glu Ala Ser Phe Glu Glu Trp Met Arg Ser Lys  
 465 470 475 480  
 Glu Val Asp Leu Asp Phe Gly Leu Thr Glu Arg Leu Arg Glu His Glu  
 485 490 495  
 15 Ala Gln Leu Met Ile Leu Ala Gln Ala Leu Asn Pro Tyr Asp Cys Leu  
 500 505 510  
 Ile His Ser Thr Pro Asn Thr Leu Val Glu Arg Gly Leu Gln Ser Ala  
 515 520 525  
 Leu Lys Tyr Glu Glu Phe Tyr Leu Lys Arg Phe Gly Gly His Tyr Met  
 20 530 535 540  
 Glu Ser Val Phe Gln Met Tyr Thr Arg Ile Ala Gly Phe Leu Ala Cys  
 545 550 555 560  
 Arg Ala Thr Arg Gly Met Arg His Ile Ala Leu Gly Arg Gln Gly Ser  
 565 570 575  
 25 Trp Trp Glu Met Phe Lys Phe Phe Phe His Arg Leu Tyr Asp His Gln  
 580 585 590  
 Ile Val Pro Ser Thr Pro Ala Met Leu Asn Leu Gly Thr Arg Asn Tyr  
 595 600 605  
 Tyr Thr Ser Ser Cys Tyr Leu Val Asn Pro Gln Ala Thr Thr Asn Gln  
 30 610 615 620  
 Ala Thr Leu Arg Ala Ile Thr Gly Asn Val Ser Ala Ile Leu Ala Arg  
 625 630 635 640  
 Asn Gly Gly Ile Gly Leu Cys Met Gln Ala Phe Asn Asp Asp Gly Thr  
 645 650 655  
 35 Ala Ser Ile Met Pro Ala Leu Lys Val Leu Asp Ser Leu Val Ala Ala  
 660 665 670  
 His Asn Lys Gln Ser Trp Thr Gly Ala Cys Val Tyr Leu Glu Pro Trp  
 675 680 685  
 His Ser Asp Val Arg Ala Val Leu Arg Met Lys Gly Val Leu Ala Gly  
 40 690 695 700  
 Glu Glu Ala Gln Arg Cys Asp Asn Ile Phe Ser Ala Leu Trp Met Pro  
 705 710 715 720  
 Asp Leu Phe Phe Lys Arg Leu Ile Arg His Leu Asp Gly Glu Glu Asn



					725					730					735	
	Val	Thr	Trp	Ser	Leu	Phe	Asp	Arg	Asp	Thr	Ser	Met	Ser	Leu	Ala	Asp
					740					745					750	
	Phe	His	Gly	Glu	Glu	Phe	Glu	Lys	Leu	Tyr	Glu	His	Leu	Glu	Ala	Met
5			755					760					765			
	Gly	Phe	Gly	Glu	Thr	Ile	Pro	Ile	Gln	Asp	Leu	Ala	Tyr	Ala	Ile	Val
			770					775					780			
	Arg	Ser	Ala	Ala	Thr	Thr	Gly	Ser	Pro	Phe	Ile	Met	Phe	Lys	Asp	Ala
	785					790					795				800	
10	Val	Asn	Arg	His	Tyr	Ile	Tyr	Asn	Thr	Gln	Gly	Ala	Ala	Ile	Ala	Gly
					805					810					815	
	Ser	Asn	Leu	Cys	Thr	Glu	Ile	Val	His	Pro	Ser	Ser	Lys	Arg	Ser	Ser
				820						825					830	
	Gly	Val	Cys	Asn	Leu	Gly	Ser	Val	Asn	Leu	Ala	Arg	Cys	Val	Ser	Arg
15			835					840					845			
	Arg	Thr	Phe	Asp	Phe	Gly	Met	Leu	Arg	Asp	Ala	Val	Gln	Ala	Cys	Val
			850				855						860			
	Leu	Met	Val	Asn	Ile	Met	Ile	Asp	Ser	Thr	Leu	Gln	Pro	Thr	Pro	Gln
	865					870					875				880	
20	Cys	Arg	His	Asp	Asn	Leu	Arg	Ser	Met	Gly	Ile	Gly	Met	Gln	Gly	Leu
					885					890					895	
	His	Thr	Ala	Cys	Leu	Lys	Met	Gly	Leu	Asp	Leu	Glu	Ser	Ala	Glu	Phe
			900						905					910		
	Arg	Asp	Leu	Asn	Thr	His	Ile	Ala	Glu	Val	Met	Leu	Leu	Ala	Ala	Met
25			915					920						925		
	Lys	Thr	Ser	Asn	Ala	Leu	Cys	Val	Arg	Gly	Ala	Arg	Pro	Phe	Ser	His
			930					935						940		
	Phe	Lys	Arg	Ser	Met	Tyr	Arg	Ala	Gly	Arg	Phe	His	Trp	Glu	Arg	Phe
	945				950					955					960	
30	Ser	Asn	Asp	Arg	Tyr	Glu	Gly	Glu	Trp	Glu	Met	Leu	Arg	Gln	Ser	Met
					965					970					975	
	Met	Lys	His	Gly	Leu	Arg	Asn	Ser	Gln	Phe	Ile	Ala	Leu	Met	Pro	Thr
			980						985					990		
	Ala	Ala	Ser	Ala	Gln	Ile	Ser	Asp	Val	Ser	Glu	Gly	Phe	Ala	Pro	Leu
35			995					1000						1005		
	Phe	Thr	Asn	Leu	Phe	Ser	Lys	Val	Thr	Arg	Asp	Gly	Glu	Thr	Leu	Arg
			1010					1015						1020		
	Pro	Asn	Thr	Leu	Leu	Leu	Lys	Glu	Leu	Glu	Arg	Thr	Phe	Gly	Gly	Lys
	1025					1030					1035				104	
40	Arg	Leu	Leu	Asp	Ala	Met	Asp	Gly	Leu	Glu	Ala	Lys	Gln	Trp	Ser	Val
					1045					1050					1055	
	Ala	Gln	Ala	Leu	Pro	Cys	Leu	Asp	Pro	Ala	His	Pro	Leu	Arg	Arg	Phe
					1060				1065					1070		

Lys Thr Ala Phe Asp Tyr Asp Gln Glu Leu Leu Ile Asp Leu Cys Ala  
 1075 1080 1085  
 Asp Arg Ala Pro Tyr Val Asp His Ser Gln Ser Met Thr Leu Tyr Val  
 1090 1095 1100  
 5 Thr Glu Lys Ala Asp Gly Thr Leu Pro Ala Ser Thr Leu Val Arg Leu  
 1105 1110 1115 112  
 Leu Val His Ala Tyr Lys Arg Gly Leu Lys Thr Gly Met Tyr Tyr Cys  
 1125 1130 1135  
 Lys Val Arg Lys Ala Thr Asn Ser Gly Val Phe Ala Gly Asp Asp Asn  
 10 1140 1145 1150  
 Ile Val Cys Thr Ser Cys Ala Leu  
 1155 1160

## (2) INFORMATION FOR SEQ ID NO:24:

15

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 269 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 20 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

25

Val Arg Arg Arg Leu Arg Cys Ala Arg Arg Arg Arg Gly Gly Pro Gly  
 1 5 10 15  
 Pro His His Asp Gln Leu Arg Arg Asp Ala Gly Arg Gly Ala Ala Gly  
 20 25 30  
 30 Pro Val Phe Arg Met Pro Ala Arg His Gly Pro His Ala Arg Val Ser  
 35 40 45  
 Pro Arg Gly His Ala Val Phe Arg Gly Ala Ser Val Val Val Thr Gln  
 50 55 60  
 Asp Glu Leu Ala Ser Val Thr Ala Val Cys Ser Gly Pro Gln Glu Ala  
 35 65 70 75 80  
 Thr His Thr Gly His Pro Gly Arg Pro Cys Ser Ala Val Thr Ile Pro  
 85 90 95  
 Ala Cys Ala Phe Val Asp Leu Asp Ala Glu Leu Cys Leu Gly Gly Pro  
 100 105 110  
 40 Gly Ala Ala Phe Leu Tyr Leu Val Phe Tyr Gln Cys Arg Asp Gln Glu  
 115 120 125  
 Leu Cys Cys Val Tyr Val Val Lys Ser Gln Leu Pro Pro Arg Gly Leu  
 130 135 140

	Glu	Ala	Ala	Leu	Glu	Arg	Leu	Phe	Gly	Arg	Leu	Arg	Ile	Thr	Asn	Thr	
	145					150				155							160
	Ile	His	Gly	Ala	Glu	Asp	Met	Thr	Pro	Leu	Pro	Pro	Asn	Arg	Asn	Val	
					165					170					175		
5	Asp	Phe	Pro	Leu	Ala	Val	Leu	Ala	Ala	Ser	Ser	Gln	Ser	Pro	Arg	Cys	
					180				185					190			
	Ser	Ala	Ser	Gln	Val	Thr	Asn	Pro	Gln	Phe	Val	Asp	Arg	Leu	Tyr	Arg	
				195				200					205				
	Trp	Gln	Pro	Asp	Leu	Arg	Gly	Arg	Pro	Thr	Ala	Arg	Thr	Cys	Thr	Tyr	
10		210					215					220					
	Ala	Ala	Phe	Ala	Glu	Leu	Gly	Val	Met	Pro	Asp	Asn	Ser	Pro	Arg	Cys	
	225					230					235					240	
	Leu	His	Arg	Thr	Glu	Arg	Phe	Gly	Ala	Val	Gly	Val	Pro	Val	Val	Ile	
					245					250						255	
15	Gly	Val	Val	Trp	Arg	Pro	Gly	Gly	Trp	Arg	Ala	Cys	Ala				
				260					265								

(2) INFORMATION FOR SEO ID NO:25:

20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 347 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

30	Met	Lys	Thr	Lys	Pro	Leu	Pro	Thr	Ala	Pro	Met	Ala	Trp	Ala	Glu	Ser
	1				5					10					15	
	Ala	Val	Glu	Thr	Thr	Thr	Ser	Pro	Arg	Glu	Leu	Ala	Gly	His	Ala	Pro
				20					25					30		
	Leu	Arg	Arg	Val	Leu	Arg	Pro	Pro	Ile	Ala	Arg	Arg	Asp	Gly	Pro	Val
35			35					40					45			
	Leu	Leu	Gly	Asp	Arg	Ala	Pro	Arg	Arg	Thr	Ala	Ser	Thr	Met	Trp	Leu
		50					55					60				
	Leu	Gly	Ile	Asp	Pro	Ala	Glu	Ser	Ser	Pro	Gly	Thr	Arg	Ala	Thr	Arg
	65					70					75					80
40	Asp	Asp	Thr	Glu	Gln	Ala	Val	Asp	Lys	Ile	Leu	Arg	Gly	Ala	Arg	Arg
					85					90					95	
	Ala	Gly	Gly	Leu	Thr	Val	Pro	Gly	Ala	Pro	Arg	Tyr	His	Leu	Thr	Arg
				100					105					110		
									204							

Gln Val Thr Leu Thr Asp Leu Cys Gln Pro Asn Ala Glu Arg Ala Gly  
 115 120 125  
 Ala Leu Leu Leu Ala Leu Arg His Pro Thr Asp Leu Pro His Leu Ala  
 130 135 140  
 5 Arg His Arg Ala Pro Pro Gly Arg Gln Thr Glu Arg Leu Ala Glu Ala  
 145 150 155 160  
 Trp Gly Gln Leu Leu Glu Ala Ser Ala Leu Gly Ser Gly Arg Ala Glu  
 165 170 175  
 Ser Gly Cys Ala Arg Ala Gly Leu Val Ser Phe Asn Phe Leu Val Ala  
 10 180 185 190  
 Ala Cys Ala Ala Ala Tyr Asp Ala Arg Asp Ala Ala Glu Ala Val Arg  
 195 200 205  
 Ala His Ile Thr Thr Asn Tyr Gly Gly Thr Arg Ala Gly Ala Arg Leu  
 210 215 220  
 15 Asp Arg Phe Ser Glu Cys Leu Arg Ala Met Val His Thr His Val Phe  
 225 230 235 240  
 Phe Val Met Arg Phe Phe Gly Gly Leu Val Ser Trp Ser His Arg Thr  
 245 250 255  
 Ser Trp Leu Asp Pro Ser Ala Ala Asp Pro Arg Arg Pro His Thr Pro  
 20 260 265 270  
 Ala Thr Arg Ala Gly Pro Val Arg Pro Leu Pro Ser Arg Pro Ala Pro  
 275 280 285  
 Leu Trp Thr Trp Thr Pro Ser Cys Ala Trp Gly Ala Leu Gly Arg Arg  
 290 295 300  
 25 Ser Cys Thr Trp Phe Ser Pro Thr Asp Ser Ala Gly Thr Arg Ser Ser  
 305 310 315 320  
 Val Ala Cys Thr Trp Ser Arg Ala Ser Ser Pro Arg Ala Asp Trp Arg  
 325 330 335  
 Arg Pro Ser Ser Gly Cys Ser Gly Ala Ser Gly  
 30 340 345

## (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 12701 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GTAGAAGTAA GGAGAGAGTG AGAGTATAGA GAAATAGATG AGAGAGAGAA GGTNAGATAA

60

	ATAAGAGGGA CAGAGAGTAG GGAGTAGGGT TAGAGAGGAT GAGGGAGAAA GATGAGAGAA	120
	GAGGGGAAGA TATATAGATG AGAGAAGGGA AGGAATGAGA GAGGGAAGAA TGATGGAAGG	180
	GAAGGAGNAG AGAAGAGTAG TAGAGGGAAA AAAGAGAAGG AGAAAAGNAA GGAAAGAAAT	240
	AGAAGAGNAG AGAAGGAAGA GAGAAAGAGA GAAGGAAGGA GGGGAAAGAA GAAANAAGAA	300
5	GGGAAGAAGG AAAGGAAGAG NAGAAGAAGG AAGAGAAGAG AAAGAAGAAG NAGAGGGGGA	360
	AGAAAAGAGA GGAAGAGGGG AAGGAGGAAA GGAGGGAGAA GAAAAANAA AAAACCACAC	420
	GGCGGCCGAA ACGTCGGGGG AACCGGTAGA AGTCCGTCAG GTCGGACGAA CCAACGGACA	480
	CCTCCGCAAA GCGCGCGCGC GCCTCCCCCG CGGCGTCGCG ACAGACCAGA TACAGCAGGG	540
	CGTGGAGGCA GTCGCGCGTG CGCGGGGGCA GCCATACCGC GTATAGGGTA ATGGCGCTGA	600
10	CGCTCTCCTC CACCCAAACG ATGCCGGGGG CTTCCATGCC ACGACGCCCC GGGGTGCGCG	660
	TGTATCGAAC GAGCGCGGCC CCAGACTTAT AGGGTGCTAA AGTTCACCGC CCCCTGCATC	720
	ATGGGCCAGG CCTCGGTGGG AAGCTCCGAC AGAGCCGCTT CGAGAATGAT GTCAGTGTTG	780
	GGCTGGGCGC CGGAGGCGTG CGTGCAGCAAG CAGCGCCCCC ACGCGGGCGC GCGCAGCTTG	840
	AAGCGCGCGC CCGCAAATC CCGCTTATGG GCCATCAGCA GCGCGTACAG CTGTCTGTGC	900
15	GTCCGGCAGG CGCTGTGGTC GATGCGGTGG GCGTCCAGCA GCTCCACGAT GGCTCGCTTG	960
	GTGAGGTTTT TAACGCGCCC CGCCCCGGGA AACGTCTGCG TGCTCTTGCG CAGCTGCACC	1020
	CCGAACAGTT CGCCCCAGAT GATCTTGAAC AGCGACAGCG CGTGCTCCGT CTCGCTCAGC	1080
	GACCCGCGCG GGGGCGAGCC GCTCAGGGCG TCGGCCACGC GCTTAACCGC GTCCTCCGAC	1140
	AGCAAGGGGC CGTCGGTCAC GTTACAGTGG CCCAGTTCGA ACACCAGCTG CATGTAGCGG	1200
20	TCGTAGTGGG GGTCAGCAG CTCCAGCACG TCCTCGGGGC TAAAGGTTTCG CCCCAGCCCC	1260
	CCGGCCATCG AGTCCCCTG CAGGCACGCG GCCATGGTGC TGCACAGACG GAACAGCTCC	1320
	CAGACGGGGG CGACGTTTAG GGTGGGGTGT AGGGCCACAA GCTCCAGCTC TCCGGCGGGC	1380
	TTGATCGTGG GGATGACGCC CGTGGCGTAG TGGTCGTAAA GCCGCCGGA GATGGCGCTG	1440
	CTATGGGCGG CCATGGGGAC GCGAAGACAG GCCTCCAGCA GCACCAGGTA GATGAACCGC	1500
25	GTGCGGCCGA CCAGGCTGTT GAGGCCGCGC ATGAGCGCGA CCACCTCGGC CGGCGCGACG	1560
	TCCGGCCGGA GGTACTTTTC GACGAAAAGG CCCACCTCCT CCGTCTCGGC GGCCTGGGCC	1620
	GACAGGGACG TGTCGGGGTC CTGGCAGCGC AGTCCCAGCA GATCCCGCTG GGCCCTCAGG	1680
	GCATCAAAAT GTATCCCCCG CAAAACAGCA CAAAAGTTCC TCGGGGTCAG CGCGGCGTCG	1740
	TGGCCCCAGA ACCGCACGTG CATGCAGTTG AGGGTCAGAA GCATGTGGAG GATGTTAAGA	1800
30	CTGTCCGCGA GGCACGCCAG CGTGCACCTC TCGAAGTAGT GCTGTACCG GAATTTGCTG	1860
	TAGATGCGCG ACCCCGCGCG CTGCGCCGCG TCGGCGTGCG ACGCGTCGCA GCGCCCTTTG	1920
	AACCGGCGGC ACAACAGGTT CGTCACCTGG GAAAAGTGTG CCGGCCACTG CCCGCTGGCG	1980
	CTCACCACGT GGTGAGCAG CATGGGCGTA AAGACGGGCT CCGAGCGCGC CCCGGACCCG	2040
	TCCATGTAGA TCAGCAGCTC CCCCTTGCGG AGAGTCCGTA CCCGCCCCAG CGACTGGTAC	2100
35	ACGGACACCA TGTCCGGCCC GTAGTTCATG GGTTCACGT AGGCGAACAT GCTGTCAAAG	2160
	TGCGGCGGAT CGAAGCTAAG GCCCACCCTC ACGACCGTTG TGTAATGAC CACCCGGTAC	2220
	CGGCCCCATG TGGTCACTTC GCCGGGCGGG GTGAGCGAGT GGAGCAGCAG CACGCGGTCC	2280
	GTAAACTGCC GGCAGAACCT GGCAACGACC TCCGCGAAGG AGACCGTCGA CAAGAAGATG	2340
	CAGACGTTAT CTCCGCCGGC CAGGCGCGCC TCCACCTCCC CGAAGAAGGT GGCCTCCGGG	2400
40	GGGCGTCCG GGGGGGGCGC CCCGCCCGCC GGCCCCGGCG GGCGCAGGGC CGCCTGCAGG	2460
	ACCTCGGGCC CCAGGCGCGG GAGAAACAGA CAACGGCGCG CCGAAAATCC GGGCATGGCA	2520
	TACTCCCCGA TGACCACGTG AACGTTCTTT TCGCCCCGGA GGCTGCACAG AAAGTCCACC	2580
	AGCTGCGCGT TGGCGGTGGC GTCCATGGCG ATGATCCGCG GGCACGTGCG CAGCAGGCGC	2640

	AGCATCAACG	CGTCGACGCG	GCCCAGCTGC	TGCATCGTCG	GCGAGTACAG	TTGGCCCCAAC	2700
	GTCGACATGA	CTTCGTCCAG	GACGAGCACG	TCGTAGTTGT	TCAACAGGTT	CGGGCCCCACG	2760
	CGATGAAGAC	TTTCCACCTG	CACGATGAGA	CGGTGGAAGG	GGCGGTTCGTT	CATGATGTAA	2820
	TTGGTGATG	AGAAGTAGGT	GACGAAGTCG	GGCAACCCTG	ACTCAGCGAA	CCGCGTCGCC	2880
5	AGGGTCTGAG	TAAAACTCCG	ACGACAGGAG	ACGACCAGCA	CACTCGTGTC	CGGAGAGTGG	2940
	ATCGCTTCCC	CCAACCAGCG	GATCAGCGCG	GTAGTTTTTC	CCGAGCCCAT	TGGCGCGCGG	3000
	ACCACAGTTA	CGCACCGGGC	CGTCGGGGCG	CTCGCGTCCG	GGAAGGTGAC	GGGTCCGTGT	3060
	TGCTGCCGCT	CGATCGTTGT	TTTCGGGTGG	ACCCGGGGAA	CCCACTCGGC	CAAATCCCCC	3120
	CCGTAAAGCA	TCCGCGCCAG	CGATACACTC	GACGTGTACT	GCTCGCACTC	GTCATCCCCG	3180
10	ATGGGACGCC	GGGCCCCCAG	GGGATCCCCC	GAGGCCGCGC	CGGGCGCCGA	CGTCGCGCCC	3240
	GGGGCGCGGG	CGGCGTGGTG	GGTCTGGTGT	GTGCAAGTGG	CGACGTTCAT	CGTCTCGGCC	3300
	ATCTGCGTCG	TGGGGCTCCT	GGTGTGCGCC	TCTGTGTTCC	GGGACAGGTT	TCCCTGCCTT	3360
	TACGCCCCCG	CGACCTCTTA	TGCGGAGGCG	AACGCCACGG	TCGAGGTGCG	CGGGGGTGTA	3420
	GCCGTCCCCC	TCCGGTTGGA	CACGCAGAGC	CTGCTGGCCA	CGTACGCAAT	TACGTCTACG	3480
15	CTGTTGCTGG	CGGCGGCCGT	GTACGCCGCG	GTGGGCGCGG	TGACCTCGCG	CTACGAGCGC	3540
	GCGCTGGATG	CGGCCCCGTC	CCTGGCGGCG	GCCCGTATGG	CGATGCCACA	CGCCACGCTA	3600
	ATCGCCGGAA	ACGTCTGCGC	GTGGCTGTTG	CAGATCACAG	TCCTGCTGCT	GGCCCCACGC	3660
	ATCAGCCAGC	TGGCCCCACT	TATCTACGTC	CTGCACTTTG	CGTGCCTCGT	GTATCTCGCG	3720
	GCCCATTTTT	GCACCAGGGG	GGTCCTGAGC	GGGACGTACC	TGCGTCAGGT	TCACGGCCTG	3780
20	ATTGACCCGG	CGCCGACGCA	CCATCGTATC	GTCGGTCCGG	TGCGGGCAGT	AATGACAAAC	3840
	GCCTTATTAC	TGGGCACCCCT	CCTGTGCACG	GCCGCCGCCG	CGGTCTCGTT	GAACACGATC	3900
	GCCGCCCTCA	ACTTCAACTT	TTCCGCCCCG	AGCATGCTCA	TCTGCCCTGAC	GACGCTGTTC	3960
	GCCCTGCTTG	TCGTGTGCTG	GTTGTTGGTG	GTCGAGGGGG	TGCTGTGTCA	CTACGTGCGC	4020
	GTGTTGGTGG	GCCCCCACCT	CGGGGCCATC	GCCGCCACCG	GCATCGTCGG	CCTGGCCTGC	4080
25	GAGCACTACC	ACACCGGTGG	CTACTACGTG	GTGGAGCAGC	AGTGGCCGGG	GGCCCAGACG	4140
	GGAGTCCGCG	TCGCCCTGGC	GCTCGTCGCC	GCCTTTGCCC	TCGCCATGGC	CGTGCTTCGG	4200
	TGCACGCGCG	CCTACCTGTA	TCACCGGCGA	CACCACACTA	AATTTTTTCGT	GCGCATGCGC	4260
	GACACCCGGC	ACCGCGCCCA	TTCGGCGCTT	CGACGCGTAC	GCAGCTCCAT	GCGCGGTTCT	4320
	AGGCGTGGCG	GGCCGCCCGG	AGACCCGGGC	TACGCGGAAA	CCCCCTACGC	GAGCGTGTCC	4380
30	CACCACGCCG	AGATCGACCG	GTATGGGGAT	TCCGACGGGG	ACCCGATCTA	CGACGAAGTG	4440
	GCCCCCGACC	ACGAGGCCGA	GCTCTACGCC	CGAGTGCAAC	GCCCCGGGCC	TGTGCCCGAC	4500
	GCCGAGCCCA	TTTACGACAC	CGTGGAGGGG	TATGCGCCAA	GGTCCGCGGG	GGAGCCGGTG	4560
	TACAGCACCG	TTCGGCGATG	GTAGCCGTTT	CGTTTCGTTT	AATAAACCGA	CGTTGTGCGT	4620
	TTCACCATAC	TTCGGCGCGC	GCGTGTGTGT	GTTTTTTTTT	TGGTGTATTAT	TTTCCCCCAC	4680
35	CCCTTCCTTT	TCTTTCGGCC	ACCACCCCCC	TCCTCCCCCG	TACTATACAA	CAAAAAATAC	4740
	CACACATACG	ACCAAATACG	GACAATCATT	TCTGTCTTTA	TTCGCTGTCA	GAGAGTGGGG	4800
	GCGTGAGCGT	GGCAGGAGGG	CGGGCCACGT	CGGGGTCCCG	CCGTCTGGTG	TGACGCGATG	4860
	GGGGGTCCGA	TGCGCGCCGG	TACTGGGGCC	CCGGCGCCCG	GGTGACCACG	CGCATGTCGG	4920
	GGGGCACGTA	GAAGTTACCC	TCTTCTTCGG	ACTCGATGTC	CACGACGTCA	AATTCGTGGG	4980
40	CGGTCAGCGA	GACGACCTCC	CCGCCGTCGG	TGATGATGAC	GTTGTGTGCG	CAGCAGCAGG	5040
	GCCGCGCCCC	GGAGAACCGG	AGGCCCATAA	CTTGGCGAGC	GTATCGTCGA	AGGCCAGGCG	5100
	GCTGTTTTGC	CGGATGTCCC	GGTATATCCC	CGGCTCGACG	CGGACGGGGG	TGATGATCAG	5160
	GGCGATCGGA	ACGGCCTGGT	CCGGGAGGAT	CGATGCCCTG	GCGGGTCCGG	GGGCCCCGCC	5220

	ACGCCCCGGCG	GGCGCTCCGC	GGCCGTCCCTC	CAGGCGGAAC	GTCACGCCCT	CCTCCGCGCC	5280
	CGCGCGGTGC	CTGCCGAGGA	ACGTCAACAG	GTGCGGTTGC	AGGGGGCAGT	CGGGAAAGTG	5340
	GCTGTGAGG	GACGTTTCCC	TGCACCAAGA	TCTGTTTGAA	GTTGCGGTGG	CGGGGGTTGG	5400
	CGAAGATGGG	CTCGCGGCGA	ACCAGCTCCC	CGGAGCTCCA	GGCCACGGA	GAGATGGTGC	5460
5	GACGCTCGAG	GTGCGGGACG	CCAAACAGAA	GCACCTCCGA	GACAACGCCG	CTATTTAACT	5520
	CCACCAGCGC	CCGATCCGGG	GCGGAGCATC	GCCTTTTTTC	GCCGGCGGCG	CGGGAATCGA	5580
	GCCAGTCCCG	GTCTTGGGTG	ACGAGCGCCT	CCTCCGGGCC	CGGGACGCGC	CCGGGCGCGA	5640
	AGTAGCGCAC	GCCGGGGTTG	GGGATGGACC	GGATGAACGC	CCGGAACGCC	TCCGGCGATC	5700
	GCCGCGCCAT	CAGGTCCCTG	TACGCGGAGG	CCGCGGGGGC	GCCGGGGTCC	GCGGGGTCGA	5760
10	ACGCGTACTT	GGCTCGGCAC	TTAACCTCGT	AGAAGGCCAG	GGGGGTCTGG	GGGGCGGGGG	5820
	CCAGGTAGCC	GTGAGGGTCC	CTGGGGCACA	CGAGGATGTC	CAGGGACGCC	CCCACCATGC	5880
	CCGTGTGGCC	GTCCATGAGG	ACCCCGCACG	CGTGACGTT	CTCCTCGGCG	AGGTCCCCGG	5940
	GTGGGTGAAA	GACGAAGCGC	CCGGCGTCGG	CGTCGTGTT	GACGCCCGCG	TCCGCGCGGC	6000
	CCACGCAGTA	GCGAAACAGC	AGGTTTCGGG	CCGTGCGCTC	GTTCAACCCG	CCGAACATCA	6060
15	CCGCCGACGA	CTGGGCGTCC	AGCCGACAGG	TGGCGTTGTG	GGTGAGCCAC	TGGGACGAGA	6120
	AGCACGGACC	CTGCGCGCCC	CACCGCAGCG	TGGAGGCGGT	CGTCAGGCC	CGCCGAAGCA	6180
	GGGCCCAGAG	CTGGCAGTCG	GCCTGGTTTT	GCGTCGCCGC	CTCGTAAAT	CCCATAAGCG	6240
	GGCGGGGGGC	GACGGCTTCG	GCGGCGGACG	GGGGGGCGCG	GCGCGTCAGG	CGCCAGAGGT	6300
	GCCGGCCGAG	CCCGCGGTCC	ACCATGCCGG	CCGCCTCCAG	CGACACGACG	AGGGAGCACA	6360
20	GATAGTCCAG	GCGAGCCAC	AGGGGCCCGA	TGGCCAGAGG	GGAGCGGACG	CCGCGCAGCA	6420
	GGCCGCGCAG	GTGGCGCTCG	AACGTTTCCG	CCAAGATATG	GGGGGGCAGT	GCGTTGGGGA	6480
	TCGCCGACGC	CGACCACATC	GGGTGCGGGT	CCGGGGGACC	GGGGCTGCAG	TCCGGGTCCG	6540
	TGGCGTGTGC	GGCCCCCGGC	GAGAGGGGAA	TGTCGGGGGT	TGGCGGGCCG	GATGAGGCCT	6600
	CAGAGAGGGC	CGGGGACGCG	GGCCGGGCCCT	TTTCGCCCCG	GGCCCCGCCG	TCGGGTTGCC	6660
25	CACGTGGGGG	GCTCTGGGGC	CAATGGGAAC	CCGGGGCCCC	CGGTGAAGTG	GGGCGGGGTG	6720
	GGGCGGGGCG	GGGCCCCAAG	ACGGTCGCCA	GATCTAGGCT	GTTGGGTCCG	GGCCGCTTCG	6780
	GGGGACTATC	GGGGTCGCGG	GCGGGGTCCG	CGGGGCGCTT	GGCGCCGGGT	GTTGCGGCGG	6840
	CCGCCATTTT	TACGAGCAGC	CGAAGAGCTC	GAGGGCGGAA	GGGATCCTCA	CGACAGAGAG	6900
	TGGCGCGCGG	CCGGGTTCGC	GTGACAGAGG	CGGGAGACCA	GCACCAGCAG	CGGCCTCAGC	6960
30	TCGGGCGGCA	GCGACACCGA	CGACAGGACG	GCCTTGTGCG	TGCGCTGGTA	ATTTATACAC	7020
	TGCTCCGTGA	ACGCGCGCCG	AATCTTGGGA	TTGCGAAGGT	GGCGCCGGAT	GCCCTCCGGC	7080
	ACGTCATACG	CCAGGCCGTG	GGTGTGGTTC	TCGGCCGAGT	TGACAAAGAG	GGCGGGGTGC	7140
	AGAACGCAGC	GATAGGCGAG	GAGGGCCACG	GCAAAGTCCG	GCGAGAGCTG	GTTGTTAAAG	7200
	TACTGGTAGC	CCGGGACGCG	GGTCACGGGG	ACGCCCAGGC	TCGGGGCCAC	GTACACGCTA	7260
35	ACCAGCAGCT	CCAGCAGCGT	CTGCCCCAGG	GCGTAGAGAT	CGACCGCCAG	CCCAGCGTCG	7320
	TGCTTCAGGG	GGCGGTTGTT	AAACTCGGCC	CGCTCGTTGT	TGAGGTACTT	TACCAAGAGC	7380
	TCCGGCGGCT	GGTTGTACCC	GTGCCCCACC	AGAGTGTGAA	AGTTGGCCGT	GGTCAGGGCG	7440
	GCGGGCATCC	CAAACCCCCG	GGGGGACTCG	AGGTCCGGCT	CCTGGAGGCA	AAACTGGCCC	7500
	CGGGATATCG	TGGAGTTGGA	GTTTCAGGGTC	ACCAGGCTAA	AGTCGGCCAG	GACGGCCCCG	7560
40	CGGAGCGACA	CCGCGTCCGA	TCGCAGCATC	ACGAGGACGT	TGGCGCACTT	GATGTCCAGG	7620
	TGGCTGATCC	CGCACCTGGT	GTTTCAGGAAC	ACCACGGCGC	GCGCCAGGTC	TGTGAAGCAG	7680
	TGGTGGAGGG	CCGTGCGGAC	GGAGGGGGTG	GTCGCGCGCA	GGGACGCCAG	CTGGCCGATG	7740
	TACTTGCCGA	GGTCCATGTC	GTACGCGGGG	AACACGATCT	GGCGCTGCTG	CAGCGAGAAC	7800

	CCGAGCGGGG	TGATAAAGCC	GCGGATGTCG	TGGGTGCGGC	CGCCGCAAAA	AGCGCACTCC	7860
	CCCACGAGCA	GGGTCGCGAC	GAGCTCCACG	GCAAACCACT	CCTTTTCCCG	GATGGTCTTC	7920
	ACGGCCAGCT	TGTGTTCGCA	AATCAACTGC	ACCTCGCCGT	ACCCCCCGA	GCCCCGAAG	7980
	CTGCGGGCCC	CGGGGATCTC	CAGGGTCGTG	TAGCGGAGGG	CGGGGTGAC	GGCGAATACG	8040
5	GGGATGCATA	GCTTGTGGAT	GCGCGCGAGG	GACAGGATGT	GCGAGGGGGG	CGACGGGGGC	8100
	GAGGTCATGG	CCGTCCTCGA	CCTGCGCAGG	GGCGGGCGCC	TCAGCTTGGC	CGCAGGGCCG	8160
	GGGGCCTCGG	GGGACGAGCG	GCGACGAGAC	GAGCGGCTCA	CTCGCCATCG	GGACAGTCCC	8220
	GCGCGAAGCC	GCTCCCGGAA	GCTGGATCGG	CGGCGGGACC	CGGGGCGGGC	TCCGGAGACG	8280
	GCGCCGTCTC	GGGGGGAGGG	GCCGCTTGGG	CGTCCGGACG	CCC GGCGGCT	GAGGGAGTGT	8340
10	ATGTAGGACG	CGAGCCAGGC	CTTGAAGGAG	CGTCGGTGTG	CACCTTGGGG	GCTGATGTCA	8400
	GCTGCCACAT	GACTAGCAGG	TCGCTGTGCG	CCGGAATCAT	CCATCCGTCC	GCCAGGTGCG	8460
	CGTCCCCCCA	CAGAGACGCG	TTCCGCGCGG	CCTCTTCGAG	CTGCTCCTCC	TGGTCCGCAA	8520
	GACGATCGTC	CGCCGCGTCC	AGGCGCTCGC	TAAGCGCGGG	ATCGAGGTAC	CGTCGGTGTG	8580
	CGGTTAGAAA	ATCACGTCGC	GCCGCTTGCT	CTTCCACGCG	AATTTTAACA	CAGGTGCGTC	8640
15	GCTGTGCGAT	CATCTCTAAG	GCGCGCGGGG	ACTTTAGCCG	CGCCTCCAAT	TCCAAGTGGG	8700
	CCGCCTTGGC	GGCCATAAAG	GCGCCAACAA	ACCTAGGATC	TTGTGTACTC	ACGCCCTCCC	8760
	GGTGTAGCTG	CAGGGTCTGG	TCCCTGTACA	CCTCGGCCCG	GAGGTGCGTC	TCGGCCAAAC	8820
	GTCGGCGCAG	GGCCGCGTGG	CTGGCGTCTC	GGCTCATCTC	GCCGCCCCCG	CGCGCGCCCCG	8880
	ACGTCGGACT	CCTTCGCCCC	GACCCCCCTG	ACCTCAGCCG	CCCCCGCCTC	GCCCGCGATG	8940
20	TTTGGCCAGC	AGCTGGCGTC	CGACGTGCAG	CAGTACCTGG	AGCGCCTGGA	GAAACAGAGG	9000
	CAACAGAAGG	TGGGCGTCGA	CGAGGCGTCG	GCGGGCCTGA	CGCTCGGCGG	CGATGCGCTG	9060
	CGCGTCCCTT	TTTTGGATTT	TGCCACCGCG	ACGCCCAAGC	GCCACCAGAC	CGTGGTCCCG	9120
	GGCGTCGGGA	CGCTCCACGA	CTGCTGCGAG	CACTCGCCGC	TCTTCTCGGC	CGTCGCGCGG	9180
	CGGTTGCTGT	TTAATAGCCT	GGTGCCGGCG	CAACTCAGGG	GGCGTGACTT	TGGGGGCGAC	9240
25	CACACGGCCA	AGCTGGAGTT	CCTGGCCCCC	GAGCTGGTGC	GGGCGGTGGC	GCGCCTGCGG	9300
	TTTCGGGAGT	GCGCGCCGGA	GGACGCCGTG	CCCCAACGCA	ACGCCTACTA	CAGCGTCTCTG	9360
	AACACGTTTC	AGGCCCTGCA	CCGCTCCGAA	GCCTTTCGGC	AGTTGGTTCA	CTTCGTGCGG	9420
	GACTTCGCCC	AGTTGTTGAA	AACCTCGTTC	CGGGCCTCTA	GTCTCGCGGA	GAATACGGGC	9480
	CCCCCGAAGA	AACGGGCCAA	GGTGGACGTG	GCCACCCACG	GGCAGACGTA	CGGCACCTTG	9540
30	GAGCTCTTCC	AGAAAATGAT	ACTAATGCAC	GCGACCTACT	TTCTGGCCGC	CGTGCTGCTC	9600
	GGGGACCACG	CGGAGCAGGT	CAACACGTTT	CTGCGGCTCG	TGTTTCGAGAT	CCCCCTGTTT	9660
	AGCGACACGG	CCGTGCGGCA	CTTCCGCCAG	CGCGCCACCG	TGTTTCTAGT	CCCCAGGCGC	9720
	CACGGAAAGA	CCTGGTTTTT	GGTGCCCCCTC	ATCGCGCTGT	CGCTCGCGTC	CTTCCGGGGG	9780
	ATCAAGATAG	GCTACACGGC	CCACATCCGC	AAGGCGACCG	AGCCCGTGTT	TGATGAGATC	9840
35	GACGCCTGCC	TGCGGGGCTG	GTTTGGCTCG	TCCCGGCTGG	ACCACGTCAA	GGGGGAAACC	9900
	ATCTCGTTCT	CGTTCCCGGA	CGGCTCGCGC	AGCACGATCG	TGTTTGCCTC	CAGCCACAAC	9960
	ACGAACGTAA	GTACGCCTTC	CTCCGCGGGT	GCCTGTTTCC	CCGGTGCCGC	CCTCCCCGAG	10020
	ATCGACCGAC	AGACAAACAC	AGCCAGACGC	GAGTGTGGGA	CGACACGCCC	GCAGCCCCCC	10080
	CCGCCATGGC	GGGGGAAGC	CTTACTGTTT	ATTTGTAATC	GGACGATGAG	GCTCTGGCCA	10140
40	CGGCCCGCGC	GACCGCGGGG	CAGCTCGTTG	CAAACAGGCG	GCTGGTATAC	GATGACAGAA	10200
	CGCAGAGGCG	CCACCCGGCG	CTGGTCGGGC	GGATGACGCT	TTCCGCGCCG	TCCCGGCCCA	10260
	CGACGACCTC	GTGCAGGTGG	GCCGTGATGC	GCGGGCGGCG	GGTCGCCTGC	CGCAGGATAA	10320
	CCGCGTCCAC	GGGGTGCCCC	AAGAGGAGCT	GACACAGGCT	CGCGTCCCCC	CGGACGGCCA	10380



	GGGTGCGCTG	GGCCATATTG	GACCACATGC	ACGGGGCGAC	GCAGGGACAG	GCCTCCGCCA	10440
	CGGCGGGGGC	GCGCCACAGC	GCGTTGGCGG	AATCGATGTG	GGCCGTCGGG	GCGCAGGCGC	10500
	CGCCTCCTCC	CGGGGGGTCTG	GTAATCCTTG	ATAGCAGCCA	TCCTAAATGG	CGGGCCCCGC	10560
	TGCCCCGGGG	ACAGAGCGAC	CCCAGGTCAT	CATCCATGGC	CCAGCAGTAT	ATGCGGCCGC	10620
5	CGGGGAGGTG	CCACCAGGCC	CCCGGACCCA	GGGCACAGCA	CGCCCCCGGA	TTCGGGGGCG	10680
	GTTCCGTGGG	TACCAGGTAG	GCGCCGTCGA	GCTCGTGGGC	CACGGGCTCG	TCCGCGAGCT	10740
	GTTCCGCGGC	GGGGTCGGGG	GTTTCCTCCG	GGGGGGAGGC	AGCTTCCAGG	TGGCCGAAGG	10800
	CTAGGGTGCA	CAGCAGCGGG	GTCCGGGGGT	GCGTTACGCT	GCGGAGGTGG	ACGGTGGCGC	10860
	AGTAGCGGCG	CTCGCGGTTA	AAGAAAAAAA	TGGCAAAAAA	CGTGTTCGAA	GGCAGGCGCA	10920
10	GCGCCTTGGG	CCGCGTCAGG	TACAGGAAAA	TCTCGCAAAA	AAGGGCACGC	TCGGGGTTCGG	10980
	GGTCCGGAAG	GGCCACCTGG	CACAGCGGCT	CGGTGAGGAC	CGTGAGGCAC	CGAAAAATCT	11040
	TAAGCCGCTC	GTCCCCCGGA	ACGACGCGCC	ACACGAAAAA	AGAGTTGGCG	ATGCGCGCGA	11100
	CAAGTCGGC	TTCGGGCCCC	GGGTCGGGGG	CGCGCGCGTC	GGGGGGGGCG	CCCCGGTGAC	11160
	CCGGCGGGGC	CGCGGCTCCC	GGGGGGCCTG	GCGTCCCCTG	GGGACGCCAC	ACTGCCCGCT	11220
15	GTGCCATGTT	GGTGGTGGGG	AAGGGACCGG	AGACGCACCA	AAAGCAGAGG	GGCCAGCGCG	11280
	TGTATGACTT	GGGGGGGGGG	GTGGGTGACC	GGTGAACAA	AAACACGCGT	CAGCGGACAA	11340
	GGCCGGGTCC	CGTACCCGCC	CCGCGACAGA	ACCGGAGTCC	GACGGCACGC	GCGACGGGGT	11400
	CTGCGAGGCT	GAGGTACGCC	GCGGTGTTAA	TGGTAAACGC	AAAGCCTCCC	GGAAAGACCA	11460
	CTAGCCCGCA	GAGGCGGCGA	TTGAACCCAA	GGCAGAGGTA	CGCGTAGCTC	TCTCCCGGAA	11520
20	GGTATTGCTC	GCAAACCTTG	TGCGGGGCAG	TGGAGGGGCT	GCCCTCCATG	AAGCGACATT	11580
	TACTCTGCTC	GCGTCCATTG	ACGTCAACGT	CAATCACCAC	TGCGATTGGA	CGGTTGGTAA	11640
	GGCGCAGCGT	GTCTCCGCTG	GTGCTGTAGT	AGTCAAACGC	GTAGTGGGCG	TCGGAGTCGG	11700
	CGAAGCGGGC	GGGGATGTCT	TCGCTGAGAG	GGACGAGCCG	CCGCCGCCGC	CCCCGACCGC	11760
	CCTGGCCGCC	CAGATGCGCC	AGCACGGCCA	GGGCGTACGC	GGTGTGAAAG	AACGCGTCGG	11820
25	GGGCGGTCCC	CTCGAGGGCG	CGCATCAGGT	TCTCCAGGAG	CACGGGGAAG	CGCCGCGTCA	11880
	CCTCCCTTAG	CCACTCGCTC	TGGTGGGGGC	CAAAGTCGTA	GCGCAGGCGC	TGGAAGATGC	11940
	GCGGGCCGCC	TTGGAGCGCG	GCCCGGATAG	AGTGGCCAG	GGCCCGCAGA	CACGCGATCT	12000
	GGATGCGCGC	GACGAAGGCC	ACCTCGGCCG	CGATGTCAAA	GGGCTGCAGC	ACGGGGCGCG	12060
	GGTGGCGCAG	GGGTCCCTCG	AGCGCGGGAA	AGCGACGCAG	CAGCGCCGTC	TGGGCCGCGG	12120
30	GGGACAGCTG	GTGGGGGCGC	ACGACGCGCT	CGGCGGCACA	GGCCTCCGTC	AGGGCCGTGG	12180
	CCAGATAGGA	GGACAACAGC	GGGGGGCGGG	TGCGTCGCCC	GCCCCACGCC	ACCGAATTTT	12240
	TGTAGGAGAC	GACGACGAAG	CGCTGCTTGG	TCCCCTAGTG	ATGGCGCAGG	ACCACGGAGA	12300
	TGGAGCGACG	GCTCCACAGC	CAGTCGGGCC	GGTCGCCGCC	GGCCAGAGGT	TCCCACCCGC	12360
	GGTCCAGCCA	CTCGACCAGC	GATCGCGGCT	TGGCGGTCCC	CGGCACGAGG	GTGAGCACGT	12420
35	CGTTGAGGAC	GTCATCGCCC	GCGGCCCGGG	GGCCCCCGGG	GGTGGCAAAG	CGCCCCCGGC	12480
	CGGGCGGTTT	CAGGCCCGCC	AGCACCGCCT	CCGCGTCCGA	CGCGCCCAGG	GCTCCCCCGC	12540
	TGACGCGNTG	GTGGACCAGG	GCGCCCTGGC	GGAGCCCCGA	GGNGACGCCG	GAGGCCGCGT	12600
	GCTTGGGGCG	CGCGCGGACC	GGGTGGCGGC	GGGTGACGTC	CTGCACGGCC	CGCTGATCAA	12660
40	GCTTGTGCGAT	ACCGTGGACT	CTGAAGTAGC	CCGTAAGGAA			12701

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 857 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

10 Met Ala Glu Thr Met Asn Val Ala Thr Cys Thr His Gln Thr His His  
     1                    5                    10                    15  
 Ala Ala Arg Ala Pro Gly Ala Thr Ser Ala Pro Gly Ala Ala Ser Gly  
                     20                    25                    30  
 Asp Pro Leu Gly Ala Arg Arg Pro Ile Gly Asp Asp Glu Cys Glu Gln  
 15                    35                    40                    45  
 Tyr Thr Ser Ser Val Ser Leu Ala Arg Met Leu Tyr Gly Gly Asp Leu  
                     50                    55                    60  
 Ala Glu Trp Val Pro Arg Val His Pro Lys Thr Thr Ile Glu Arg Gln  
     65                    70                    75                    80  
 20 Gln His Gly Pro Val Thr Phe Pro Asp Ala Ser Ala Pro Thr Ala Arg  
                     85                    90                    95  
 Cys Val Thr Val Val Arg Ala Pro Met Gly Ser Gly Lys Thr Thr Ala  
                     100                    105                    110  
 Leu Ile Arg Trp Leu Gly Glu Ala Ile His Ser Pro Asp Thr Ser Val  
 25                    115                    120                    125  
 Leu Val Val Ser Cys Arg Arg Ser Phe Thr Gln Thr Leu Ala Thr Arg  
                     130                    135                    140  
 Phe Ala Glu Ser Gly Leu Pro Asp Phe Val Thr Tyr Phe Ser Ser Thr  
     145                    150                    155                    160  
 30 Asn Tyr Ile Met Asn Asp Arg Pro Phe His Arg Leu Ile Val Gln Val  
                     165                    170                    175  
 Glu Ser Leu His Arg Val Gly Pro Asn Leu Leu Asn Asn Tyr Asp Val  
                     180                    185                    190  
 Leu Val Leu Asp Glu Val Met Ser Thr Leu Gly Gln Lys Pro Thr Met  
 35                    195                    200                    205  
 Gln Gln Leu Gly Arg Val Asp Ala Leu Met Leu Arg Leu Leu Arg Thr  
                     210                    215                    220  
 Cys Pro Arg Ile Ile Ala Met Asp Ala Thr Ala Asn Ala Gln Leu Val  
     225                    230                    235                    240  
 40 Asp Phe Leu Cys Ser Leu Arg Gly Glu Lys Asn Val His Val Val Ile  
                     245                    250                    255  
 Gly Glu Tyr Ala Met Pro Gly Phe Ser Ala Arg Arg Cys Leu Phe Leu  
                     260                    265                    270

Pro Arg Leu Gly Pro Glu Val Leu Gln Ala Ala Leu Arg Pro Pro Gly  
 275 280 285  
 Pro Ala Gly Gly Ala Pro Pro Pro Asp Ala Pro Pro Asp Ala Thr Phe  
 290 295 300  
 5 Phe Gly Glu Val Glu Ala Arg Leu Ala Gly Gly Asp Asn Val Cys Ile  
 305 310 315 320  
 Phe Leu Ser Thr Val Ser Phe Ala Glu Val Val Ala Arg Phe Cys Arg  
 325 330 335  
 10 Gln Phe Thr Asp Arg Val Leu Leu Leu His Ser Leu Thr Pro Pro Gly  
 340 345 350  
 Glu Val Thr Thr Trp Gly Arg Tyr Arg Val Val Ile Tyr Thr Thr Val  
 355 360 365  
 Val Thr Val Gly Leu Ser Phe Asp Pro Pro His Phe Asp Ser Met Phe  
 370 375 380  
 15 Ala Tyr Val Lys Pro Met Asn Tyr Gly Pro Asp Met Val Ser Val Tyr  
 385 390 395 400  
 Gln Ser Leu Gly Arg Val Arg Thr Leu Arg Lys Gly Glu Leu Leu Ile  
 405 410 415  
 Tyr Met Asp Gly Ser Gly Ala Arg Ser Glu Pro Val Phe Thr Pro Met  
 20 420 425 430  
 Leu Leu Asn His Val Val Ser Ala Ser Gly Gln Trp Pro Ala Gln Phe  
 435 440 445  
 Ser Gln Val Thr Asn Leu Leu Cys Arg Arg Phe Lys Gly Arg Cys Asp  
 450 455 460  
 25 Ala Ser His Ala Asp Ala Ala Gln Arg Ser Arg Ile Tyr Ser Lys Phe  
 465 470 475 480  
 Arg Tyr Lys His Tyr Phe Glu Arg Cys Thr Leu Ala Cys Leu Ala Asp  
 485 490 495  
 Ser Leu Asn Ile Leu His Met Leu Leu Thr Leu Asn Cys Met His Val  
 30 500 505 510  
 Arg Phe Trp Gly His Asp Ala Ala Leu Thr Pro Arg Asn Phe Cys Leu  
 515 520 525  
 Phe Leu Arg Gly Ile His Phe Asp Ala Leu Arg Ala Gln Arg Asp Leu  
 530 535 540  
 35 Arg Glu Leu Arg Cys Gln Asp Pro Asp Thr Ser Leu Ser Ala Gln Ala  
 545 550 555 560  
 Ala Glu Thr Glu Glu Val Gly Leu Phe Val Glu Lys Tyr Leu Arg Pro  
 565 570 575  
 Asp Val Ala Pro Ala Glu Val Val Met Arg Gln Ser Leu Val Gly Arg  
 40 580 585 590  
 Thr Arg Phe Ile Tyr Leu Val Leu Leu Glu Ala Cys Leu Arg Val Pro  
 595 600 605  
 Met Ala Ala His Ser Ser Ala Ile Phe Arg Arg Leu Tyr Asp His Tyr

	610		615		620	
	Ala Thr Gly Val Ile Pro Thr Ile Asn Ala Ala Gly Glu Leu Glu Leu					
	625		630		635	640
	Val His Pro Thr Leu Asn Val Ala Pro Val Trp Glu Leu Phe Arg Leu					
5		645		650		655
	Cys Ser Thr Met Ala Ala Cys Leu Gln Trp Asp Ser Met Ala Gly Gly					
		660		665		670
	Ser Gly Arg Thr Phe Ser Pro Glu Asp Val Leu Glu Leu Leu Asn Pro					
	675		680		685	
10	His Tyr Asp Arg Tyr Met Gln Leu Val Phe Glu Leu Gly His Cys Asn					
	690		695		700	
	Val Thr Asp Gly Pro Leu Leu Ser Glu Asp Ala Val Lys Arg Val Ala					
	705		710		715	720
	Asp Ala Leu Ser Gly Cys Pro Pro Arg Gly Ser Val Ser Glu Thr Glu					
15		725		730		735
	His Ala Leu Ser Leu Phe Lys Ile Ile Trp Gly Glu Leu Phe Gly Val					
		740		745		750
	Gln Leu Ala Lys Ser Thr Gln Thr Phe Pro Gly Ala Gly Arg Val Lys					
	755		760		765	
20	Asn Leu Thr Lys Arg Ala Ile Val Glu Leu Leu Asp Ala His Arg Ile					
	770		775		780	
	Asp His Ser Ala Cys Arg Thr Gln Leu Tyr Ala Leu Leu Met Ala His					
	785		790		795	800
	Lys Arg Glu Phe Ala Gly Ala Arg Phe Lys Leu Arg Ala Pro Ala Trp					
25		805		810		815
	Gly Arg Cys Leu Arg Thr His Ala Ser Gly Ala Gln Pro Asn Thr Asp					
		820		825		830
	Ile Ile Ala Ala Leu Ser Glu Leu Pro Thr Glu Ala Trp Pro Met Met					
	835		840		845	
30	Gln Gly Ala Val Asn Phe Ser Thr Leu					
	850		855			

## (2) INFORMATION FOR SEQ ID NO:28:

## 35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 470 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Val Tyr Cys Ser His Ser Ser Ser Pro Met Gly Arg Arg Ala Pro Arg  
 1 5 10 15  
 Gly Ser Pro Glu Ala Ala Pro Gly Ala Asp Val Ala Pro Gly Ala Arg  
 5 20 25 30  
 Ala Ala Trp Trp Val Trp Cys Val Gln Val Ala Thr Phe Ile Val Ser  
 35 40 45  
 Ala Ile Cys Val Val Gly Leu Val Leu Ala Ser Val Phe Arg Asp  
 50 55 60  
 10 Arg Phe Pro Cys Leu Tyr Ala Pro Ala Thr Ser Tyr Ala Glu Ala Asn  
 65 70 75 80  
 Ala Thr Val Glu Val Arg Gly Gly Val Ala Val Pro Leu Arg Leu Asp  
 85 90 95  
 Thr Gln Ser Leu Leu Ala Thr Tyr Ala Ile Thr Ser Thr Leu Leu  
 15 100 105 110  
 Ala Ala Ala Val Tyr Ala Ala Val Gly Ala Val Thr Ser Arg Tyr Glu  
 115 120 125  
 Arg Ala Leu Asp Ala Ala Arg Arg Leu Ala Ala Ala Arg Met Ala Met  
 130 135 140  
 20 Pro His Ala Thr Leu Ile Ala Gly Asn Val Cys Ala Trp Leu Leu Gln  
 145 150 155 160  
 Ile Thr Val Leu Leu Leu Ala His Arg Ile Ser Gln Leu Ala His Leu  
 165 170 175  
 Ile Tyr Val Leu His Phe Ala Cys Leu Val Tyr Leu Ala Ala His Phe  
 25 180 185 190  
 Cys Thr Arg Gly Val Leu Ser Gly Thr Tyr Leu Arg Gln Val His Gly  
 195 200 205  
 Leu Ile Asp Pro Ala Pro Thr His His Arg Ile Val Gly Pro Val Arg  
 210 215 220  
 30 Ala Val Met Thr Asn Ala Leu Leu Leu Gly Thr Leu Leu Cys Thr Ala  
 225 230 235 240  
 Ala Ala Ala Val Ser Leu Asn Thr Ile Ala Ala Leu Asn Phe Asn Phe  
 245 250 255  
 Ser Ala Pro Ser Met Leu Ile Cys Leu Thr Thr Leu Phe Ala Leu Leu  
 35 260 265 270  
 Val Val Ser Leu Leu Leu Val Val Glu Gly Val Leu Cys His Tyr Val  
 275 280 285  
 Arg Val Leu Val Gly Pro His Leu Gly Ala Ile Ala Ala Thr Gly Ile  
 290 295 300  
 40 Val Gly Leu Ala Cys Glu His Tyr His Thr Gly Gly Tyr Tyr Val Val  
 305 310 315 320  
 Glu Gln Gln Trp Pro Gly Ala Gln Thr Gly Val Arg Val Val Ala Ala  
 325 330 335

Phe Ala Met Ala Val Leu Arg Cys Thr Arg Ala Tyr Leu Tyr His Arg  
                   340                          345                          350  
 Arg His His Thr Lys Phe Phe Val Arg Met Arg Asp Thr Arg His Arg  
                   355                          360                          365  
 5 Ala His Ser Ala Leu Arg Arg Val Arg Ser Ser Met Arg Gly Ser Arg  
                   370                          375                          380  
 Arg Gly Gly Pro Pro Gly Asp Pro Gly Tyr Ala Glu Thr Pro Tyr Ala  
 385                          390                          395                          400  
 Ser Val Ser His His Ala Glu Ile Asp Arg Tyr Gly Asp Ser Asp Gly  
 10                          405                          410                          415  
 Asp Pro Ile Tyr Asp Glu Val Ala Pro Asp His Glu Ala Glu Leu Tyr  
                   420                          425                          430  
 Ala Arg Val Gln Arg Pro Gly Pro Val Pro Asp Ala Glu Pro Ile Tyr  
                   435                          440                          445  
 15 Asp Thr Val Glu Gly Tyr Ala Pro Arg Ser Ala Gly Glu Pro Val Tyr  
                   450                          455                          460  
 Ser Thr Val Arg Arg Trp  
 465                          470

20 (2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 687 amino acids  
 (B) TYPE: amino acid  
 25 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met Ala Ala Ala Ala Thr Pro Gly Ala Lys Arg Pro Ala Asp Pro Ala  
 1                          5                          10                          15  
 Arg Asp Pro Asp Ser Pro Pro Lys Arg Pro Arg Pro Asn Ser Leu Asp  
 35                          20                          25                          30  
 Leu Ala Thr Val Phe Gly Pro Arg Pro Ala Pro Pro Arg Pro Thr Ser  
                   35                          40                          45  
 Pro Gly Ala Pro Gly Ser His Trp Pro Gln Ser Pro Pro Arg Gly Gln  
                   50                          55                          60  
 40 Pro Asp Gly Gly Ala Pro Gly Glu Lys Ala Arg Pro Asp Ala Leu Ser  
                   65                          70                          75                          80  
 Glu Ala Ser Ser Gly Pro Pro Thr Pro Asp Ile Pro Leu Ser Pro Gly  
                           85                          90                          95

Gly Ala His Ala Ile Asp Pro Asp Cys Ser Pro Gly Pro Pro Asp Pro  
 100 105 110  
 Asp Pro Met Trp Ser Ala Ser Ala Ile Pro Asn Ala Leu Pro Pro His  
 115 120 125  
 5 Ile Leu Ala Glu Thr Phe Glu Arg His Leu Arg Gly Leu Leu Arg Gly  
 130 135 140  
 Val Arg Ser Pro Leu Ala Ile Gly Pro Leu Trp Ala Arg Leu Asp Tyr  
 145 150 155 160  
 Leu Cys Ser Leu Val Val Ser Leu Glu Ala Ala Gly Met Val Asp Arg  
 10 165 170 175  
 Gly Leu Gly Arg His Leu Trp Arg Leu Thr Arg Arg Ala Pro Pro Ser  
 180 185 190  
 Ala Ala Glu Ala Val Ala Pro Arg Pro Leu Met Gly Phe Tyr Glu Ala  
 195 200 205  
 15 Ala Thr Gln Asn Gln Ala Asp Cys Gln Leu Trp Ala Leu Leu Arg Arg  
 210 215 220  
 Gly Leu Thr Thr Ala Ser Thr Leu Arg Trp Gly Ala Gln Gly Pro Cys  
 225 230 235 240  
 Phe Ser Ser Gln Trp Leu Thr His Asn Ala Ser Leu Arg Leu Asp Ala  
 20 245 250 255  
 Gln Ser Ser Ala Val Met Phe Gly Arg Val Asn Glu Pro Thr Ala Arg  
 260 265 270  
 Asn Leu Leu Phe Arg Tyr Cys Val Gly Arg Ala Asp Ala Gly Val Asn  
 275 280 285  
 25 Asp Asp Ala Asp Ala Gly Arg Phe Val Phe His Gln Pro Gly Asp Leu  
 290 295 300  
 Ala Glu Glu Asn Val His Ala Cys Gly Val Leu Met Asp Gly His Thr  
 305 310 315 320  
 Gly Met Val Gly Ala Ser Leu Asp Ile Leu Val Cys Pro Arg Asp Pro  
 30 325 330 335  
 His Gly Tyr Leu Ala Pro Ala Pro Gln Thr Pro Leu Ala Phe Tyr Glu  
 340 345 350  
 Val Lys Cys Arg Ala Lys Tyr Ala Phe Asp Pro Ala Asp Pro Gly Ala  
 355 360 365  
 35 Pro Ala Ala Ser Ala Tyr Glu Asp Leu Met Ala Arg Arg Ser Pro Glu  
 370 375 380  
 Ala Phe Arg Ala Phe Ile Arg Ser Ile Pro Asn Pro Gly Val Arg Tyr  
 385 390 395 400  
 Phe Ala Pro Gly Arg Val Pro Gly Pro Glu Glu Ala Leu Val Thr Gln  
 40 405 410 415  
 Asp Arg Asp Trp Leu Asp Ser Arg Ala Ala Gly Glu Lys Arg Arg Cys  
 420 425 430  
 Ser Ala Pro Asp Arg Ala Leu Val Glu Leu Asn Ser Gly Val Val Ser

	435		440		445	
	Glu Val Leu Leu Phe Gly Val Pro Asp Leu Glu Arg Arg Thr Ile Ser					
	450		455		460	
5	Pro Val Ala Trp Ser Ser Gly Glu Leu Val Arg Arg Glu Pro Ile Phe					
	465		470		475	480
	Ala Asn Pro Arg His Pro Asn Phe Lys Gln Ile Leu Val Gln Gly Asn					
		485		490		495
	Val Pro Arg Gln Pro Leu Ser Arg Leu Pro Pro Ala Thr Ala Pro Gly					
		500		505		510
10	Asp Val Pro Arg Gln Ala Pro Arg Gly Arg Gly Gly Gly Arg Asp Val					
		515		520		525
	Pro Pro Gly Gly Arg Pro Arg Ser Ala Arg Arg Ala Trp Arg Gly Pro					
		530		535		540
	Arg Thr Arg Gln Gly Ile Asp Pro Pro Gly Pro Gly Arg Ser Asp Arg					
15	545		550		555	560
	Pro Asp His His Pro Arg Pro Arg Arg Ala Gly Asp Ile Pro Gly His					
		565		570		575
	Pro Ala Lys Gln Pro Pro Gly Leu Arg Arg Tyr Ala Arg Gln Val Met					
		580		585		590
20	Gly Leu Ala Phe Ser Gly Ala Arg Pro Cys Cys Cys Arg His Asn Val					
		595		600		605
	Ile Ile Thr Asp Gly Gly Glu Val Val Ser Leu Thr Ala His Glu Phe					
		610		615		620
	Asp Val Val Asp Ile Glu Ser Glu Glu Glu Gly Asn Phe Tyr Val Pro					
25	625		630		635	640
	Pro Asp Met Arg Val Val Thr Arg Ala Pro Gly Pro Gln Tyr Arg Arg					
		645		650		655
	Ala Ser Asp Pro Pro Ser Arg His Thr Arg Arg Arg Asp Pro Asp Val					
		660		665		670
30	Ala Arg Pro Pro Ala Thr Leu Thr Pro Pro Leu Ser Asp Ser Glu					
		675		680		685

## (2) INFORMATION FOR SEQ ID NO:30:

## 35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 107 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:



Val Thr Phe Leu Gly Arg His Arg Ala Gly Ala Glu Glu Gly Val Thr  
 1 5 10 15  
 Phe Arg Leu Glu Asp Gly Arg Gly Ala Pro Ala Gly Arg Gly Gly Ala  
 5 20 25 30  
 Pro Gly Pro Ala Lys Ala Ser Ile Leu Pro Asp Gln Ala Val Pro Ile  
 35 40 45  
 Ala Leu Ile Ile Thr Pro Val Arg Val Glu Pro Gly Ile Tyr Arg Asp  
 50 55 60  
 10 Ile Arg Arg Asn Ser Arg Leu Ala Phe Asp Asp Thr Leu Ala Lys Leu  
 65 70 75 80  
 Trp Ala Ser Arg Ser Pro Gly Arg Gly Pro Ala Ala Ala Asp Thr Thr  
 85 90 95  
 Ser Ser Ser Pro Thr Ala Gly Arg Ser Ser Arg  
 15 100 105

## (2) INFORMATION FOR SEQ ID NO:31:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 525 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 25 (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Val Gly Gly Arg Arg Pro Gly Gly Arg Met Asp Glu Ser Gly Arg Gln  
 30 1 5 10 15  
 Arg Pro Ala Ser His Val Ala Ala Asp Ile Ser Pro Gln Gly Ala His  
 20 25 30  
 Arg Arg Ser Phe Lys Ala Trp Leu Ala Ser Tyr Ile His Ser Leu Ser  
 35 35 40 45  
 Arg Arg Ala Ser Gly Arg Pro Ser Gly Pro Ser Pro Arg Asp Gly Ala  
 50 55 60  
 Val Ser Gly Ala Arg Pro Gly Ser Arg Arg Arg Ser Ser Phe Arg Glu  
 65 70 75 80  
 Arg Leu Arg Ala Gly Leu Ser Arg Trp Arg Val Ser Arg Ser Ser Arg  
 40 85 90 95  
 Arg Arg Ser Ser Pro Glu Ala Pro Gly Pro Ala Ala Lys Leu Arg Arg  
 100 105 110  
 Pro Pro Leu Arg Arg Ser Glu Thr Ala Met Thr Ser Pro Pro Ser Pro

	115		120		125
	Pro Ser His Ile Leu Ser Leu Ala Arg Ile His Lys Leu Cys Ile Pro				
	130		135		140
	Val Phe Ala Val Asn Pro Ala Leu Arg Tyr Thr Thr Leu Glu Ile Pro				
5	145		150		155
	Gly Ala Arg Ser Phe Gly Gly Ser Gly Gly Tyr Gly Glu Val Gln Leu				
		165		170	175
	Ile Cys Glu His Lys Leu Ala Val Lys Thr Ile Arg Glu Lys Glu Trp				
		180		185	190
10	Phe Ala Val Glu Leu Val Ala Thr Leu Leu Val Gly Glu Cys Ala Phe				
		195		200	205
	Cys Gly Gly Arg Thr His Asp Ile Arg Gly Phe Ile Thr Pro Leu Gly				
		210		215	220
	Phe Ser Leu Gln Gln Arg Gln Ile Val Phe Pro Ala Tyr Asp Met Asp				
15	225		230		235
	Leu Gly Lys Tyr Ile Gly Gln Leu Ala Ser Leu Arg Ala Thr Thr Pro				
		245		250	255
	Ser Val Ala Thr Ala Leu His His Cys Phe Thr Asp Leu Ala Arg Ala				
		260		265	270
20	Val Val Phe Leu Asn Thr Arg Cys Gly Ile Ser His Leu Asp Ile Lys				
		275		280	285
	Cys Ala Asn Val Leu Val Met Leu Arg Ser Asp Ala Val Ser Leu Arg				
		290		295	300
	Arg Ala Val Leu Ala Asp Phe Ser Leu Val Thr Leu Asn Ser Asn Ser				
25	305		310		315
	Thr Ile Ser Arg Gly Gln Phe Cys Leu Gln Glu Pro Asp Leu Glu Ser				
		325		330	335
	Pro Arg Gly Phe Gly Met Pro Ala Ala Leu Thr Thr Ala Asn Phe His				
		340		345	350
30	Thr Leu Val Gly His Gly Tyr Asn Gln Pro Pro Glu Leu Leu Val Lys				
		355		360	365
	Tyr Leu Asn Asn Glu Arg Ala Glu Phe Asn Asn Arg Pro Leu Lys His				
		370		375	380
	Asp Val Gly Leu Ala Val Asp Leu Tyr Ala Leu Gly Gln Thr Leu Leu				
35	385		390		395
	Glu Leu Leu Val Ser Val Tyr Val Ala Pro Ser Leu Gly Val Pro Val				
		405		410	415
	Thr Arg Val Pro Gly Tyr Gln Tyr Phe Asn Asn Gln Leu Ser Pro Asp				
		420		425	430
40	Phe Ala Val Leu Ala Tyr Arg Cys Val Leu His Pro Ala Leu Phe Val				
		435		440	445
	Asn Ser Ala Glu Thr Asn Thr His Gly Leu Ala Tyr Asp Val Pro Glu				
		450		455	460

Gly Ile Arg Arg His Leu Arg Asn Pro Lys Ile Arg Arg Ala Phe Thr  
 465 470 475 480  
 Glu Gln Cys Ile Asn Tyr Gln Arg Thr His Lys Ala Val Leu Ser Ser  
 485 490 495  
 5 Val Ser Leu Pro Pro Glu Leu Arg Pro Leu Leu Val Leu Val Ser Arg  
 500 505 510  
 Leu Cys His Ala Asn Pro Ala Ala Arg His Ser Leu Ser  
 515 520 525

10 (2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 79 amino acids

(B) TYPE: amino acid

15 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Met Ala Val Ser Asp Leu Arg Arg Gly Gly Arg Leu Ser Leu Ala Ala  
 1 5 10 15  
 Gly Pro Gly Ala Ser Gly Asp Glu Arg Arg Arg Asp Glu Arg Leu Thr  
 25 20 25 30  
 Arg His Arg Asp Ser Pro Ala Arg Ser Arg Ser Arg Lys Leu Asp Arg  
 35 40 45  
 Arg Arg Asp Pro Gly Arg Ala Pro Glu Thr Ala Pro Ser Arg Gly Glu  
 50 55 60  
 30 Gly Pro Leu Gly Arg Pro Asp Ala Arg Arg Leu Arg Glu Cys Met  
 65 70 75

(2) INFORMATION FOR SEQ ID NO:33:

35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 217 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

40 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Met Ser Arg Asp Ala Ser His Ala Ala Leu Arg Arg Arg Leu Ala Glu  
 1 5 10 15  
 Thr His Leu Arg Ala Glu Val Tyr Arg Asp Gln Thr Leu Gln Leu His  
 5 20 25 30  
 Arg Glu Gly Val Ser Thr Gln Asp Pro Arg Phe Val Gly Ala Phe Met  
 35 40 45  
 Ala Ala Lys Ala Ala His Leu Glu Leu Glu Ala Arg Leu Lys Ser Arg  
 50 55 60  
 10 Ala Arg Leu Glu Met Met Arg Gln Arg Ala Thr Cys Val Lys Ile Arg  
 65 70 75 80  
 Val Glu Glu Gln Ala Ala Arg Arg Asp Phe Leu Thr Ala His Arg Arg  
 85 90 95  
 Tyr Leu Asp Pro Ala Leu Ser Leu Asp Ala Ala Asp Asp Arg Leu Ala  
 15 100 105 110  
 Asp Gln Glu Glu Gln Leu Glu Glu Ala Ala Ala Asn Ala Ser Leu Trp  
 115 120 125  
 Gly Asp Gly Asp Leu Ala Asp Gly Trp Met Ser Pro Gly Asp Ser Asp  
 130 135 140  
 20 Leu Leu Val Met Trp Gln Leu Thr Ser Ala Pro Lys Val His Thr Asp  
 145 150 155 160  
 Ala Pro Ser Arg Pro Gly Ser Arg Pro Thr Tyr Thr Pro Ser Ala Ala  
 165 170 175  
 Gly Arg Pro Asp Ala Gln Ala Ala Pro Pro Pro Glu Thr Ala Pro Ser  
 25 180 185 190  
 Pro Glu Pro Ala Pro Gly Pro Ala Ala Asp Pro Ala Ser Gly Ser Gly  
 195 200 205  
 Phe Ala Arg Asp Cys Pro Asp Gly Glu  
 210 215

30

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 493 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Val Tyr Ser Arg Pro Pro Gly Val Ala Ala Gly Ser Gly Pro Cys Thr

	1	5	10	15
	Pro Arg Pro Gly Gly Ala Ser Arg Pro Asn Val Gly Ala Gly Pro Arg			
	20	25	30	
5	Gly Trp Arg Leu Gly Ser Ser Arg Arg Pro Arg Ala Arg Pro Thr Ser			
	35	40	45	
	Asp Ser Phe Ala Pro Thr Pro Leu Thr Ser Ala Ala Pro Asp Ala Met			
	50	55	60	
	Phe Gly Gln Gln Leu Ala Ser Asp Val Gln Gln Tyr Leu Glu Arg Leu			
	65	70	75	80
10	Glu Lys Gln Arg Gln Gln Lys Val Gly Val Asp Glu Ala Ser Ala Gly			
	85	90	95	
	Leu Thr Leu Gly Gly Asp Ala Leu Arg Val Pro Phe Leu Asp Phe Ala			
	100	105	110	
	Thr Ala Thr Pro Lys Arg His Gln Thr Val Val Pro Gly Val Gly Thr			
15	115	120	125	
	Leu His Asp Cys Cys Glu His Ser Pro Leu Phe Ser Ala Val Ala Arg			
	130	135	140	
	Arg Leu Leu Phe Asn Ser Leu Val Pro Ala Gln Leu Arg Gly Arg Asp			
	145	150	155	160
20	Phe Gly Gly Asp His Thr Ala Lys Leu Glu Phe Leu Ala Pro Glu Leu			
	165	170	175	
	Val Arg Ala Val Ala Arg Leu Arg Phe Arg Glu Cys Ala Pro Glu Asp			
	180	185	190	
	Ala Val Pro Gln Arg Asn Ala Tyr Tyr Ser Val Leu Asn Thr Phe Gln			
25	195	200	205	
	Ala Leu His Arg Ser Glu Ala Phe Arg Gln Leu Val His Phe Val Arg			
	210	215	220	
	Asp Phe Ala Gln Leu Leu Lys Thr Ser Phe Arg Ala Ser Ser Leu Ala			
	225	230	235	240
30	Glu Asn Thr Gly Pro Pro Lys Lys Arg Ala Lys Val Asp Val Ala Thr			
	245	250	255	
	His Gly Gln Thr Tyr Gly Thr Leu Glu Leu Phe Gln Lys Met Ile Leu			
	260	265	270	
	Met His Ala Thr Tyr Phe Leu Ala Ala Val Leu Leu Gly Asp His Ala			
35	275	280	285	
	Glu Gln Val Asn Thr Phe Leu Arg Leu Val Phe Glu Ile Pro Leu Phe			
	290	295	300	
	Ser Asp Thr Ala Val Arg His Phe Arg Gln Arg Ala Thr Val Phe Leu			
	305	310	315	320
40	Val Pro Arg Arg His Gly Lys Thr Trp Phe Leu Val Pro Leu Ile Ala			
	325	330	335	
	Leu Ser Leu Ala Ser Phe Arg Gly Ile Lys Ile Gly Tyr Thr Ala His			
	340	345	350	

Ile Arg Lys Ala Thr Glu Pro Val Phe Asp Glu Ile Asp Ala Cys Leu  
 355 360 365  
 Arg Gly Trp Phe Gly Ser Ser Arg Val Asp His Val Lys Gly Glu Thr  
 370 375 380  
 5 Ile Ser Phe Ser Phe Pro Asp Gly Ser Arg Ser Thr Ile Val Phe Ala  
 385 390 395 400  
 Ser Ser His Asn Thr Asn Val Ser Thr Pro Ser Ser Arg Gly Ala Cys  
 405 410 415  
 Phe Pro Gly Ala Ala Leu Pro Glu Ile Asp Arg Gln Thr Asn Thr Ala  
 10 420 425 430  
 Arg Arg Glu Cys Gly Thr Trp Gln Pro Pro Pro Pro Trp Arg Gly Glu  
 435 440 445  
 Ala Leu Leu Phe Ile Cys Asn Arg Thr Met Arg Leu Trp Pro Arg Pro  
 450 455 460  
 15 Ala Arg Pro Arg Gly Ser Ser Leu Gln Thr Gly Gly Trp Tyr Thr Met  
 465 470 475 480  
 Thr Glu Arg Arg Gly Ala Thr Arg Arg Trp Ser Gly Gly  
 485 490

20 (2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 399 amino acids  
 (B) TYPE: amino acid  
 25 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Val Phe Leu Phe His Arg Ser Pro Thr Pro Pro Pro Lys Ser Tyr Thr  
 1 5 10 15  
 Arg Trp Pro Leu Cys Phe Trp Cys Val Ser Gly Pro Phe Pro Thr Thr  
 35 20 25 30  
 Asn Met Ala Gln Arg Ala Val Trp Arg Pro Gln Gly Thr Pro Gly Pro  
 35 40 45  
 Pro Gly Ala Ala Ala Pro Pro Gly His Arg Gly Ala Pro Pro Asp Ala  
 50 55 60  
 40 Arg Ala Pro Asp Pro Gly Pro Glu Ala Asp Leu Val Ala Arg Ile Ala  
 65 70 75 80  
 Asn Ser Val Phe Val Trp Arg Val Val Arg Gly Asp Glu Arg Leu Lys  
 85 90 95

Ile Phe Arg Cys Leu Thr Val Leu Thr Glu Pro Leu Cys Gln Val Pro  
 100 105 110  
 Asp Pro Asp Pro Glu Arg Ala Leu Phe Cys Glu Ile Phe Leu Tyr Leu  
 115 120 125  
 5 Trp Lys Ala Leu Arg Leu Pro Ser Asn Thr Phe Phe Ala Ile Phe Phe  
 130 135 140  
 Phe Asn Arg Glu Arg Arg Tyr Cys Ala Thr Val His Leu Arg Ser Val  
 145 150 155 160  
 Thr His Pro Arg Thr Pro Leu Leu Cys Thr Leu Ala Phe Gly His Leu  
 10 165 170 175  
 Glu Ala Asp Pro Glu Glu Thr Pro Asp Pro Ala Ala Glu Gln Leu Ala  
 180 185 190  
 Asp Glu Pro Val Ala His Glu Leu Asp Gly Ala Tyr Leu Val Pro Thr  
 195 200 205  
 15 Glu Pro Pro Pro Asn Pro Gly Ala Cys Cys Ala Leu Gly Pro Gly Ala  
 210 215 220  
 Trp Trp His Leu Pro Gly Gly Arg Ile Tyr Cys Trp Ala Met Asp Asp  
 225 230 235 240  
 Asp Leu Gly Ser Leu Cys Pro Pro Gly Ser Arg Ala Arg His Leu Gly  
 20 245 250 255  
 Trp Leu Leu Ser Arg Ile Thr Asp Pro Pro Gly Gly Gly Gly Ala Cys  
 260 265 270  
 Ala Pro Thr Ala His Ile Asp Ser Ala Asn Ala Leu Trp Arg Ala Pro  
 275 280 285  
 25 Ala Val Ala Glu Ala Cys Pro Cys Val Ala Pro Cys Met Trp Ser Asn  
 290 295 300  
 Met Ala Gln Arg Thr Leu Ala Val Arg Gly Asp Ala Ser Leu Cys Gln  
 305 310 315 320  
 Leu Leu Phe Gly His Pro Val Asp Ala Val Ile Leu Arg Gln Ala Thr  
 30 325 330 335  
 Arg Arg Pro Arg Ile Thr Ala His Leu His Glu Val Val Val Gly Arg  
 340 345 350  
 Asp Gly Ala Glu Ser Val Ile Arg Pro Thr Ser Ala Gly Trp Arg Leu  
 355 360 365  
 35 Cys Val Leu Ser Ser Tyr Thr Ser Arg Leu Phe Ala Thr Ser Cys Pro  
 370 375 380  
 Ala Val Ala Arg Ala Val Ala Arg Ala Ser Ser Ser Asp Tyr Lys  
 385 390 395

40

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 452 amino acids

(B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```

Phe Leu Thr Gly Tyr Phe Arg Val His Gly Ile Asp Lys Leu Asp Gln
10 1      5      10      15
Arg Ala Val Gln Asp Val Thr Arg Arg His Pro Val Arg Ala Arg Pro
      20      25      30
Lys His Ala Ala Ser Gly Val Xaa Ser Gly Leu Arg Gln Gly Ala Leu
      35      40      45
15 Val His Xaa Ala Val Ser Gly Gly Ala Leu Gly Ala Ser Asp Ala Glu
      50      55      60
Ala Val Leu Ala Gly Leu Glu Pro Pro Gly Gly Gly Arg Phe Ala Thr
65      70      75      80
Pro Gly Gly Pro Arg Ala Ala Gly Asp Asp Val Leu Asn Asp Val Leu
20      85      90      95
Thr Leu Val Pro Gly Thr Ala Lys Pro Arg Ser Leu Val Glu Trp Leu
      100      105      110
Asp Arg Gly Trp Glu Pro Leu Ala Gly Gly Asp Arg Pro Asp Trp Leu
      115      120      125
25 Trp Ser Arg Arg Ser Ile Ser Val Val Leu Arg His His Tyr Gly Thr
      130      135      140
Lys Gln Arg Phe Val Val Val Ser Tyr Lys Asn Ser Val Ala Trp Gly
145      150      155      160
Gly Arg Arg Trp Pro Leu Leu Ser Ser Tyr Leu Ala Thr Ala Leu Thr
30      165      170      175
Glu Ala Cys Ala Ala Glu Arg Val Val Arg Pro His Gln Leu Ser Pro
      180      185      190
Ala Ala Gln Thr Ala Leu Leu Arg Arg Phe Pro Ala Leu Glu Gly Pro
      195      200      205
35 Leu Arg His Pro Arg Pro Val Leu Gln Pro Phe Asp Ile Ala Ala Glu
      210      215      220
Val Ala Phe Val Ala Arg Ile Gln Ile Ala Cys Leu Arg Ala Leu Gly
225      230      235      240
His Ser Ile Arg Ala Ala Leu Gln Gly Gly Pro Arg Ile Phe Gln Arg
40      245      250      255
Leu Arg Tyr Asp Phe Gly Pro His Gln Ser Glu Trp Leu Gly Glu Val
      260      265      270
Thr Arg Arg Phe Pro Val Leu Leu Glu Asn Leu Met Arg Ala Leu Glu
225

```



275                      280                      285  
 Gly Thr Ala Pro Asp Ala Phe Phe His Thr Ala Tyr Ala Val Leu Ala  
 290                      295                      300  
 His Leu Gly Gly Gln Gly Gly Arg Gly Arg Arg Arg Arg Leu Val Pro  
 5 305                      310                      315                      320  
 Leu Ser Asp Asp Ile Pro Ala Arg Phe Ala Asp Ser Asp Ala His Tyr  
 325                      330                      335  
 Ala Phe Asp Tyr Tyr Ser Thr Ser Gly Asp Thr Leu Arg Leu Thr Asn  
 340                      345                      350  
 10 Arg Pro Ile Ala Val Val Ile Asp Gly Asp Val Asn Gly Arg Glu Gln  
 355                      360                      365  
 Ser Lys Cys Arg Phe Met Glu Gly Ser Pro Ser Thr Ala Pro His Arg  
 370                      375                      380  
 Val Cys Glu Gln Tyr Leu Pro Gly Glu Ser Tyr Ala Tyr Leu Cys Leu  
 15 385                      390                      395                      400  
 Gly Phe Asn Arg Arg Leu Cys Gly Leu Val Val Phe Pro Gly Gly Phe  
 405                      410                      415  
 Ala Phe Thr Ile Asn Thr Ala Ala Tyr Leu Ser Leu Ala Asp Pro Val  
 420                      425                      430  
 20 Ala Arg Ala Val Gly Leu Arg Phe Cys Arg Gly Ala Gly Thr Gly Pro  
 435                      440                      445  
 Gly Leu Val Arg  
 450

25 (2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26339 base pairs  
 (B) TYPE: nucleic acid  
 30 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

35 GGGAGCGAGG TGAGGGAATG AGAGAAGAAA GGAGGAGGAT AGTAAGTGGG GAGATGAAGA 60  
 AGAAAATAGA CGCTGGGAAG GAGAATACGA CGAGGAGGGA GGAAGGAGAA GAGACTCGGA 120  
 AAGCATAAGA GGTGGAAGG AGGTGTGAGT AATTGAGCGG AGATGAGAGG ATAGGATAAG 180  
 GCGAGAGACG GAAGATAAGA AAGTGAAGGG AGTAAGGGTA AGGTGAGAGA GAGAAAGAGG 240  
 40 AGGATTATGT GATGGTTAGG GGAGAGAGAG GAATATGTGG AGAAATTGTG AGGAAGGAAA 300  
 AGAGAGAGAA GAGTGTGGGT ATAAAGGAGA TATGGATGGA ATAGAGTAAA TTGAAGAAGG 360  
 GAAGAGATGG TAAGAATGGA GTGAAGAGGA GGTAAAGAAT TTAGTAAAGG AGTGGTGTATG 420  
 ATGGATAAAA AAGTGGAAT GGGGTAAGAA AGAAGAGAGA GGAGGAGGAA AAAAAAAAAA 480

	TAAAAAAGG	GCCTTGCCGC	CGGCCCTGGC	GTACGCGCTA	TATAAGCCCA	TGCGGTATTG	540
	GATGAGTTCC	CGCGCGCCCC	GGAATCCTC	CACCGCCAC	GGGGCCAGGT	CCGCGGCCGC	600
	CGCGTCAAAC	TCCGCCAGCA	GGCGCCCCAG	GGCGTCAAAG	TTCATCTCCC	AGGGCACCCCT	660
	GCGCACCAACC	TCATCCCGCA	GCCGGGCGCA	CAGGGCGGTG	TGCTTGGTGA	CGCGCGCGCC	720
5	CAGCTCCTCC	ACGGCCTCCG	CGCGCTCGGC	GCCCTTGGCG	CCCAGGACGC	CCTGGTACCT	780
	GGCGGAAAGG	CGCTCGTAGG	CCGGCTGGGC	CCGCAGCCCC	GACACCGTGT	TGGTGGTGTG	840
	CTGCAGGGCG	CGCAGCTGCT	CGTGCATGGC	GCGGAACCCC	TCGGGGGACT	TCCAGGCGCC	900
	CCCCCGGACG	CGGCCAAAGC	GACCCAGAC	CTCGTCCCAC	TCCGCCTCGG	CCTCCTCCAG	960
	GGACCTCCGC	AGGGCGTCGA	CGCGGCGCCG	AGTATCAAAG	AGCGCCCCCA	GGCGGCCGGC	1020
10	GTGCCGCGCC	AGGGGGCCGG	GGCCGTCGCC	GCGGGCGGCG	CTTAGCGGGT	GCGTCTCGAA	1080
	GGTGCGCTGG	GCGTGCTCTA	GCCAGATAAC	CGCGGGCACG	TCGAGCTCGC	GCGTTTTCTC	1140
	GGTCTGATCC	AACAGAACCCT	CGACCTGGTC	GGCGATCTCC	GCCACCGAGC	GCGCCTGGTC	1200
	GAGCGTCTTG	GCCACGGTCG	CCGGGACGGC	GACCACCTTC	AGCATGGTCT	TGAGGTTGGC	1260
	CAGGCCCTCG	GCCTCGATCT	GGGCCCGGCG	CTCGCGCGCG	GCCAGCGCCT	CCCGCAGGCC	1320
15	CGCCATGACC	CGCTCGGTGG	CCTCCGCGCG	CTGCTGTTTG	GCGCGCACCA	CTGCGTCTTT	1380
	GGTCTCGGCC	GTGTCTTGCC	GGGTACGAA	GGCGACATAC	TCGGCGTACG	CCGTGTTCTT	1440
	CACGGGGCTC	TGGTCCACGC	GCTCCAACGC	CGCCGCGCAC	GCGACCAGCG	CGTCTCGCT	1500
	GGGACACGGC	AGGGTGACCC	CGGTCCGGAC	CAGCTCCGCG	GTGGCCTCCG	GGTCATTCCG	1560
	GGCCGCGGAT	ATCTGCTCCG	CGGCGGCCGC	CAGGTCCAGG	GGCAGCCCGC	CGAGCGCCCC	1620
20	GTGCACGTCG	GCCCGGATGG	CGTCCAGGCG	ATCGCGGAGC	TCCACGTAGT	CGGCGTAGCC	1680
	ATGTTGGAAG	AACGGCACGT	ACCGGCGCAG	GCCGGGCACG	CTCGTCATGT	CGTCCGCCAG	1740
	GCGCCCCACG	GCCTCGTGGT	AGTCGATAAA	CCCGTCGCCC	GCCTGGGCCA	TTTCCAGGAG	1800
	CCCCTCGCG	ATGCGCAGCA	GCCGCGCCAG	GGGCTCGGCG	TCGACCCGAA	ACATGTCGGC	1860
	GTAGGTTTCG	GCGGCGGCGT	GGAACGCCGC	GCTCCAGCCG	AGGCGGTGGA	TGGCGGCGAG	1920
25	CGGGGGGAGC	ATGGGGTGGC	GCTGGTTCTC	GGGGGTGTAG	GGGTAAACG	CGAAGGCCGT	1980
	ATCCAGGGCG	AGGGTGACCG	CCTCGGCGTT	GGCCGCGAGC	GCCTGCTCGG	CGCGCTTGCG	2040
	GAAGTCCCGG	GGGTTGTAGC	CGTGCCTGCC	CGCCAGCGCC	TGCAGGCGGC	GCAGCTCGAC	2100
	CACGTCTGAAC	TGGGCGCGGT	TCTCGACGCG	GTCAGCGGCC	GCCTCGACGC	CGGCGGCCCA	2160
	GCGCTCGCTG	CTGCCCCGGG	CGCGCTGGGC	CGCCATCTTC	CCCGTCAGGT	CGGCGACGGC	2220
30	GGCCTCAAGT	TCCTCGGCGC	GGCGTCGCGT	GGCGCCGATG	ACCTTGCCCA	GCTCCTGCAG	2280
	GGCGCGCCCG	CTGGGGGAAT	GGTCCCCGGC	CGTCCCTTCG	GCGTGCAGCA	GGCCCCCGAA	2340
	CCCAGCCTCG	TGCCCCGCGA	GGCTTTCCCG	AGCAGCGGTG	GTCGCGCGGG	CCGCGGCATC	2400
	GATGAGGGCG	GCATGGTCCC	CCTCCGCTG	GGCGCAGGCC	CGGCGCGCAT	GGAATACCAG	2460
	GTCGGCGGCC	GCCGACCCCA	GGGTGGTGAG	CTTGTCGATG	GCCCCCGCG	CCTCCAGGGC	2520
35	CAGCCGAGTC	GCCTTTACAT	ACCCCGCGGC	GCTATCGGCC	AGCGCCGCGA	GGAAGGACAG	2580
	GGGCGAGGCC	GGGTGCGGGG	CGGCCCGGCC	CAGGGCCGAC	ACCGCGTCCG	CCAGGGCGCC	2640
	ATGCGCCCGC	ACGGCCGCGT	CCACCGTCGC	CGCGGGACTT	GCCGTCGCGA	CGGCGGCGCT	2700
	CCCGGCGTTG	ATGGCGTTTG	ACACGGCTTT	GGCGATTGTG	GGGGCGTGAT	CGGAAAAGAA	2760
	CTGCACGAGG	ACCGGCGTCT	CGGGGCGGTC	NGCGAACATG	GTCTTCAGCA	CCACCACTAA	2820
40	GGCGGGATGC	ATGCCGGCCA	CAACCGTCTC	GGTATCCGGG	GTCTGGTGTT	CCANGGCCCTC	2880
	CCGGTACTGC	CCCATCACCC	CCCACATGTC	CGCCCGCAGC	CCCGCCTTGA	CTTCCGGGGG	2940
	GGGGCCCCCG	GACGGCATCG	GCCAGGTCGG	TCCACCCCGC	GGGGCAGGGA	GGCCCGCAGG	3000
	GTGCGCAGCA	CGGCCGGACA	CGCCTTTAGC	CCCACAAAGT	CCGGGAGGGG	CCGCAGGACC	3060

	CCTTGGAGTT	TGTGCAGGAA	CTTCTCCCGG	GCGTCGTGGG	CCACCTTGGC	GCGCTCCCGC	3120
	GCGTCGTTGA	GCATCGCCTC	CAGGGCGTGG	GCGCGCTCCC	GAAGCCGGGA	GCGCGCCTCC	3180
	GGAGCGAGCT	CCGCCGTCAT	CTTGGCCGCC	TCCATGGCCC	TCGCCTGCCG	CAGCGCGTCT	3240
	TCGGCCATGC	GCGTGGCCTC	GGGGGACAGC	CCGCCCCCGT	CGACGTACGG	CGCGGGGCCG	3300
5	GTCGCCGGGA	CGAAGGCCGC	GTCGCTGTCC	AGCTGCTGCG	CGAGCGCCGC	GTCGAGGGCG	3360
	TCGAAGCGCT	GCAGTTCGGC	CAGCCCCGAG	CTGCGCCGCG	CCTGCTGGTC	GTTGATGCCG	3420
	TGGATGCTGC	GCGCCAGCTC	TTCAGGGGC	TTGCGTTCGA	TGAGCCCTTG	GGTCGCGGCG	3480
	TCGGTCAGGA	CCGAGAGCCA	GGCCGCCAGG	TCCTCGGGGG	CATCTAGGGT	CTGGCCCCGC	3540
	TGGAGCAGGT	CCCGCAGCAG	GATGGCCTGG	GGGCTGGTGG	CGAGGGGGGG	CGGGGGGGGG	3600
10	AGCGCGCGCG	GCTGAGCGAC	GTCCCGCGTG	TGTTGGTCAA	AGGCCGCTAG	CGATTCCAGC	3660
	AACTGGACCA	TGGGCACGAC	CGCGGCCGAG	GCCACGTGAA	ACCGACAGTC	GTGGCTGTCT	3720
	CTGGCCTGCA	GGGCCTTCGC	GCTGTATACG	GCTCCCCGGT	GGAAGTACTC	CTTGATCGCG	3780
	CTTTCGATCG	CCCGCGGGG	CTGGATCCGC	ACGTCTTCCA	GCCGCGCATG	GATGGCCTCG	3840
	GGGCCCAGGG	CGGGCGGGCA	CGGGGCCCTG	CCGCCGGCGC	CCCGGGGGGG	GGGGGCAACG	3900
15	GGCATCACGG	TCAGGGGGCC	CGGCGCGCTG	CGAGACCGAG	TAGACCCCGC	GGGCGAGGGC	3960
	GTATAAGGCC	TCGCGCATCT	CGCGGGCCTC	CGCCTCGACC	CGCATCTTTT	TGCCCCGGGC	4020
	AAAATGGGCC	AGCGCCTGGA	TCCGATGGAG	AAGCGGCTCC	GGGTGCGTCG	GGGTGGCGGG	4080
	GGCGAACAGG	GTGTTCGGGT	GGGCGCGCGA	GCGCTCCAGG	AGCCACTTTC	CGAGGCGTGC	4140
	GTACAGATTG	GCCGGCGGGG	CGGCGCGCAG	CTGCAGATCC	AGGTCCGCGA	GGTCCCCGTA	4200
20	AAAGGCGTCC	GTCTCCCGAA	TAAAGTCCCT	GGCGACCAGG	ACCAGCTTAG	CGAGGGCCAG	4260
	GCGCCCGATC	TGCGAATTTT	TGTCCAGCAC	GTGCTGGATG	AGGGGCCGGT	GGGCGGCCAC	4320
	GTCCGCCAGG	CTCATGCGCG	TGGACGCCAG	GAAGTCCCCG	ACGGCCGTTT	TGCGGGGCGG	4380
	CATGCCGAGG	GTGAAGTCCA	GCAGGGCCGC	GGCCGGGCCG	GCCACCCCGG	CCTGCGTATG	4440
	CGTGCGGGCC	CCGTCTCGA	TCAAAAAGGC	GAGGACGCGC	TCAAAGAAGA	AGATGACGCA	4500
25	GAGCTCCAAC	AGCCCCGGGT	GCGCCGGGTA	CGGCGACCGC	AGGGCGTTGA	TGGTGAGCTG	4560
	CGAACACGCG	GCCACCTCGC	GGGCCAGGGC	GGCATCGCGC	GCCGCGAGCC	GGACCGCCGT	4620
	GGCGGCCACA	TTGGGGTGA	CCTCGAACAG	CTGCGCCAGG	TCGGCGCCGG	GGGGCTCCGG	4680
	GGGGCGGCGG	GCCCCAGCG	TCTCGAGCAC	GGACGGCGAC	GACGGGCTCG	CGGGCCCGTC	4740
	ATCGCCGCCT	CCCTGCCCGG	ACTGCGGGGG	GGTATCCGGT	GCGGGAGGGA	CCGTGGCGGC	4800
30	TATGGGCGTC	GGGAGGAGG	CGGGGACCTC	GCGCGCGACG	GGGGCCTTCT	TCTTGGGCGC	4860
	GGACTTCTTC	TTGGCCTTGG	CGGGCGGGGC	CTTGGGGGCG	GGCCTCTCGC	CCGAGGTCAG	4920
	ATCCTCCACG	CTGGACGGTG	GGGTCCAGGT	GGGCCGCGCG	CGCTTGGGCA	AGCCGGTAGA	4980
	ATAGCGCGCC	CGGTGGCAAC	CCACCGGCAC	TGCCCCCACC	TCCAGGACCC	GCAGGTCCTC	5040
	GGCTTCTTCG	GCCGCGTCCC	CGGCGGGTGT	CTGCGGGGGC	GGGGCGGCGT	GCGGTGGACC	5100
35	CGAGGCCGCG	GCGTCCGGGG	CCGAGGGCTT	TGCGGGCGGG	GTCCCCCTCA	GGGCTGCTGC	5160
	CCACACATCA	TCGGGGGGGC	GGTTTGGGTG	CCCCGCCTGC	GGTGTGTTGG	GTGGGCCCCGA	5220
	GGCCCCCGG	GGGGCCTCGG	GGGGCCGGTC	GGCCCGAGGG	GTCTGGACGT	GGGTGGGCGC	5280
	GGGGAGCGCG	GGGACGACCG	GGCCCGAGCC	TTCTCCGTCC	CCCTTGGGGA	CCACACCGAC	5340
	AAAGAGCGCC	CCTAGCCCCC	CGATCTCGCC	CCGCAGGGGG	TGGGTGATGG	CCACGCGCCG	5400
40	CTCGACGAAC	GGTTCGTCTT	GCAGGTAAGT	CTCGCTGGCC	CCGTAGAGGT	GCAGGGCCGC	5460
	GGCGGTCAGG	TCCGCCGGCG	CCACGGCCCC	CGGGCCGGAG	GGCACAAAAA	ACACCATGGC	5520
	GCCCCCCAC	CGCACCTTGG	GGCGGTCGTG	GGCGTAATAC	GTCAGGTACG	GGTACACGTC	5580
	GCCCCCGCG	ACCTTGGCGA	TAAACGCGGG	CGTTCCTCGG	GGCAGGCCGT	GCGGGTCAAA	5640

	CAGATAGGCC	GTGTCGCCGT	CCCGGTAGAG	CCCCATGCCC	AGGGGGCCGA	TGGTCAGGAG	5700
	CGTGTAGGAC	AGCGGCCGCA	TGGCCCAGGG	GCCGGCGAAG	AACGTGTGCG	CGGGGCATTG	5760
	CGTCTCCAGC	AGCCCCGCCG	TGGGCTCCCC	GAAGAAGCCC	ACCTCGCCGT	ACACCCGCGA	5820
	AAACACGCAA	CGCAGGCCGC	CGCGCGCCGC	CGGGTACTCC	AGGAAGTTGG	GGAGCTCGAT	5880
5	AATGGAACAC	ATGCGCGGCG	GCCCGGAGCC	CGCGGCGCGC	CGCGTCCACT	CGCCCCCTC	5940
	CACCAGACAT	CCCTCAATGG	CCTCCGCGGA	CAGCACGTGC	CGGGGCCCCA	CGTCGAAAAG	6000
	AAGACTGAGA	AACGACAGGG	ACGAGCGCAT	GCACGATACC	GACCCCCCCG	GCTCCAGATC	6060
	GGTCGCGAAC	TGGTTCCGAA	CACCGGTGAC	CACGATATCG	CGATCCCCCT	GGGCGCTTCA	6120
	TCGTGGGGTG	AGGTAGCGCG	GCCGGAATCA	TGTGTGCCGC	GCCCCCCACG	AGCGGGGCCT	6180
10	GTTTATGGGC	CGGGCGTCCC	GATGAGTACT	GTTGTTTCCG	CCGCCCGAAC	CCCCCCGCC	6240
	CATCAACCGC	CTGTTCTGTC	CCCTAACCAC	ACACCCGGTA	TCGCGTGAGC	TCGTCTGAAC	6300
	TGAACAGGAG	CACGCGGGCG	CAGGTCGCCC	ACGGGCCCCA	CGCCAGGCGC	AGCGCCGCAA	6360
	CCGTGTACGG	GTCGTACACG	CCTTGGGCGT	CGCACGCGAC	CGGCAGGGAG	ACGAACAGCC	6420
	CGCCCCGCGT	GGGGACGCGC	GGCAGGAGGT	CCGGGTGCGC	CGGGATGACG	GGGGCTAGGA	6480
15	TCGCCCCCAC	CGCATCCGCC	GGCACGTAGG	CGGCAAACGC	CGAACGCCAC	GGGGTGCACT	6540
	CGCCGGGCGC	GTGGGGCCGG	GTCTGGGTTT	CGACCCGGA	GTTCGCGGCC	GCCCCGCCGT	6600
	CGGGGCGGCC	GCGCACGAGG	GCGGACAGCG	GGACCCCGC	CGCCGCCAGG	CACTCGCTGG	6660
	AGATGATGAC	GTGAATCAGC	GAGGCGGGGC	TGCTCGGGTC	CCGGGTGAGA	TCGTATTGGA	6720
	CCTCGTTGGC	AAAGTGCGCG	TTCATGGCCC	GGCCGCGGGT	GCGAGCCCTT	CCCAGTGCCG	6780
20	GAAGGGGCGT	GGGTGGGGGG	TGCGTGTGCG	CGTCTCGGG	GCCCCGCGGC	GCACGTGCGC	6840
	TTATACGCTG	TGTGTTTCGT	CTGTCCCCAG	GGAATCCGGG	GCCAGGACTT	TAACCTGCTT	6900
	TTCGTCGACG	AGGCCAACTT	TATTCGCCCG	GATGCGGTCC	AGACGATTAT	GGGCTTTCTC	6960
	AATCAGGCCA	ACTGCAAGAT	CATCTTCGTC	TCGTGACCA	ACACCGGGAA	GGCCAGCACG	7020
	AGCTTTTGT	ACAACCTCCG	CGGGGCCGCC	GACGAGCTGC	TCAACGTGGT	CACCTATATA	7080
25	TGCGACGACC	ACATGCCGCG	GGTGGTGACG	CACACCAACG	CCACGGCCTG	TTCTTGCTAT	7140
	ATCCTGAACA	AACCCGTGTT	TATCACGATG	GACGGCGCCG	TTCCGCCGAC	GGCCGATCTG	7200
	TTTCTGCCCC	ACTCCTTCAT	GCAGGAGATC	ATCGGGGGGC	AGGCCCGCGA	GACCGGCGAC	7260
	GACCGGCCCC	TCCTAACAAA	GTCGGCGGGG	GAGCGGTTTC	TGCTGTACCG	CCCCTCCACC	7320
	ACCACCAACA	GCGGCCTGAT	GGCCCCCGAG	CTGTACGTGT	ACGTGGACCC	GGCGTTCACG	7380
30	GCCAACACGC	GCGCCTCCGG	CACCGGCATC	GCGGTCTGCG	GGAGGTACCG	CGACGATTTC	7440
	ATTATCTTCG	CCCTGGAGCA	CTTTTTCCTC	CGCGCGCTCA	CGGGATCGGC	CCCCGCGGAC	7500
	ATCGCCCGCT	GCGTCGTGCA	CAGCCTCGCC	CAGGTGCTGG	CGCTGCACCC	CGGGGCGTTT	7560
	CGCAGCGTTC	GCGTGGCGGT	CGAGGGCAAC	AGCAGCCAGG	ACTCGGCCGT	GGCCATCGCC	7620
	ACACACGTGC	ATACCGAGAT	GCACCGCATC	CTGGCCTCGG	CGGGGGCCAA	CGGCCCGGGG	7680
35	CCCGAGCTCC	TCTTCTATCA	CTGCGAGCCG	CCCGCGGGCG	CGGTATTGTA	CCCCTTCTTT	7740
	CTGCTCAACA	AACAGAAGAC	GCCCGCCTTC	GAATACTTTA	TCAAAAAGTT	CAACTCCGGG	7800
	GGCGTCATGG	CGTCCAGGA	GCTCGTCTCC	GTGACGGTGC	GCCTGCAGAC	CGACCCGGTC	7860
	GAGTATCTGT	CCGAGCAGCT	CAACAACCTC	ATCGAAAACG	TCTCTCCCAA	CACCGACGTC	7920
	CGCATGTACT	CCGGAAAACG	CAACGGTGCC	GCGGACGACC	TCATGGTCGC	GGTCATCATG	7980
40	GCCATTTACC	TGGCGGCCCC	GACCGGGATC	CCCCCGGCCT	TTTTTCCGAT	CACGCGCACG	8040
	TCTTGAGTCT	TTCTTGCCGT	TTCTTTTGTT	TCTCTTTCTT	TCCCCCTCT	CTCCGCAATA	8100
	AACGCCTTCC	CGGAACGTG	TTTTCCCCC	CTACAACAGT	GTGTCCGTT	GGTTGGGTGG	8160
	TTGGGGTGCG	GGGGTGGGCG	GGGAAGCAA	GAAACGGTC	GGCGAACACA	ACATCGGGAA	8220

	AACGGATTCC	CGAACGTGCG	TCTTCCCAGA	TTGACACAC	ACCCCCCTTC	TCCTTAAATA	8280
	AACACAAACC	ACACGCTCGT	TGGTTGGTTA	ATGCCGGCGC	TTTATTTACG	TCTTGTTTTT	8340
	TTGCGTTTCC	TCCGCGGGTC	CCTTCCCAAC	ACGCTGCCCC	CCGCCTCAGG	GGTAGCGGAT	8400
	AACCGGGGCC	ATGTCGCCCG	ATTGCACAAC	GGCGGCGCCG	TCGAACGTAC	ACACCCGAAC	8460
5	CGCCGGGGCC	AGGGCCAGGA	TGTCCCCGAG	TTGGCCCGCG	TGCGCCAGCC	AGGCGACCAG	8520
	CGCCTCGTAA	AGCGGCAGCC	TGCGCTCGCC	GTCCTGCATC	AGCATGGGGG	CTTCGGGGTG	8580
	GATGAGCTGG	GCGGCTTCTC	GCGTGACGCT	CTGCATCTGC	AGGAGCGCGT	TCACGTATCC	8640
	GTCCTGGGCG	CTCAGCGCGA	GCAGCCGGGG	GATGAGCGTG	AGGATGAGGG	TGGTTCCTTC	8700
	GGTTATGGAG	TAGACCATGT	TGAGGACGAG	CGACCGCAGC	TCGGTGTTTA	CGGAGGCGAG	8760
10	TTGCTGGACG	TCGGCCACGA	GCGAGAGACG	GGCCCCGTTG	TAATACAGCA	CGTTGAGGTC	8820
	GGGGAGCTCC	CCGGGCGTCC	GGGGGTCGGG	GTTGAGGTCC	CGGATGCCCC	GGGCGACCAG	8880
	CCGCGCGACT	ATCTCGCGGG	CCAGGGGCGT	TGGGAGCGGG	ACCGGAAACC	GCAGCGTGAG	8940
	GTCCAGCGAC	TCCAGGCGCA	CGTCCGTCGC	CTGGCCCTCG	AAGACGGGCG	GGACGAGGCT	9000
	GACGGGATCC	CCGTTGCAGA	GGTCGACGGG	GGAGGTGTTG	CGGAGATTGA	CGGTGCCGGC	9060
15	GTGCGTGAGC	CCCAGGTCCA	CGGGGCAGGC	GACGATTTCG	GTGGGCAGCA	CCCGCGTGAT	9120
	TACCGCGGGG	AAGCGCCTGC	GGTACGCCAG	CAACAACCCC	AACGTGTCCG	GACTAACTCC	9180
	TCCGGAGACG	AACGATTTCG	GCGCCACGTC	CGCGAGCGCC	AGCTGGCGGC	GGATGGTCCG	9240
	CAGAAAGACC	ACTCGACCCT	CGCACCCTG	CAGCGCCGCG	GCATCGGGGC	GCGAGATACC	9300
	CGAGGGGATC	GCGATGTCTG	CTTCGAAACA	ATCCGTGATC	ATGGCGCCCG	GCCGCGAGAC	9360
20	ACCGGAACGC	GGGGGTGCGG	GAGGGCCGGA	AAGCGCAACG	CAACCGGGAC	GATGATGAAA	9420
	CAGAGATGGG	GGGCACCGAC	CGTGTGGGAG	AGGGGGCGGG	GCAGGGCTCA	GCAGCACGCA	9480
	CGGGGAGGTC	TGTCGTGCGC	AGGAGCCCCA	GGTGAGAATC	AGTCCCCCGG	AGCTCGGGTC	9540
	TGGGTTTTAT	TGGGACCTGC	CCTCGGAATC	GCGCTCCCA	GTCCAAGCCC	CCCCGGGGGG	9600
	GCGGGGACAG	GGGGTGTGTG	TGGGTAAAAG	CAACGTCCGA	AAATCAAACC	CAATGCCCCA	9660
25	AACAGGAAAA	AAAAAAAAGA	CGGGCGGGTG	GAGGGAAAGC	TGGGGAAGAA	GAAGCCAATT	9720
	TTACAGAGAC	AGGCCCTTTA	GCGGGGAGGC	GTCGTAGATG	AGATACTGCG	TAAAGTGGGT	9780
	CTCTCGCGCG	TGGGCCTCCC	CATCGCGGGC	GCTGCGTAGC	AGGGCGGGGT	CGCTGGCGCA	9840
	GGTGATCGGG	TAGGCTTCCT	GAAACAGGCC	GCACGGGTCT	TCCACGAGCT	CGCGGCACCC	9900
	CGGCGGGCGC	TTAAACTGCA	CGTCGCTGGC	AGCGGTGGCC	GTGGATACCG	CCGATCCCCT	9960
30	TTCCACGATA	AGACGCTCCA	GGCAGCGATG	TTTGGCCGTG	ATGTCGGCCG	CGGTGAAGAA	10020
	CTTGAAGCAG	GGGCTGAGGA	CGGGCGAGGC	CCCGTTGAGG	TGATAGGCCC	CGTTGTACAG	10080
	CAGGTCCCCG	TACGAGAACC	GCTGCGACGC	CCACGGGTTG	GCCGTGGCCG	CGAAGGGCCG	10140
	CGCCGGGTCG	CTCTGGCCGT	GGTCGTACAT	GAGGGCTATG	ACGTCCCCCT	CCTTGTCCCC	10200
	CGCGTACACG	CCGCCGGCCG	CGCGTCCCCG	CGGGTTGCAG	GGCCGGCGAA	AGTAGTTGAT	10260
35	GTCCGTGGCC	ACGGGGGTGG	CGATGAACTC	ACACACGGCA	TCCTGCCCCG	GGTCCATGCC	10320
	GGCGCGCCGC	GGCACCTGGG	CGCAGCCAAA	GACCGGGAGG	GGCTGGGCCG	GCCCCAGCCG	10380
	GTTTCCCGCC	ACGACCGCGT	TGCGCAGGTA	CACGGCGGCC	GCGTTGTTTA	GCAGCGGGGG	10440
	GGCCCCGCGG	CCGAGGTAAA	AGTTTGGGGG	GAGGTTGCCC	ATGTCCGTAA	CGGGGTGCG	10500
	GACGGTGGCC	GTGGCCGCGA	CGGCGGTGTA	GCCCACACCC	AGGTCCACGT	TTCCGCGCGG	10560
40	CTGGGTGAGC	GTGAAGCTGA	CCCCCCCCGC	CGTTTCGTGG	CGGGCCACCT	GGAGCTGGCC	10620
	CAGAAAGTAC	GCCTCCGACG	CGCGCTCGGA	AAACAGCACG	TTTTCGGTCA	CGAAGCGGTC	10680
	CTGCCGCACG	ACGGTGAACC	CGAACCCGGG	GTTGAGGCCC	GTCTTGAGCT	GGTGATACAG	10740
	GGCCACGGGG	CTCATCTTGA	AGTACCCCGC	CATGAGCGCG	TAGGTCAGCG	CGTTCTCCCC	10800

	CGCCGCGCTC	TCGCGGGCGT	GCTGCACCAC	GGGCTGGCGG	ATGGAGGAGA	AGTAGTTGGC	10860
	CCCCAGGGCC	GGGGGACCA	GGGGGACGTG	GCGCGCCAGG	TCGCGCAGGG	CCGGGGGGAA	10920
	GTTGGGCGCG	TTGGCCACGT	GGTCGGCGCC	CGCAAACAGC	GCGTGGACGG	GCAGGACGTA	10980
	GAAGTATTCG	CCATTTTGGA	TGGTGTGGTC	CAGGTGCTGG	GGGGCCATGA	GCAGCACGCC	11040
5	GGCGTGCAGC	GCCCCGTCAA	AAATGCGCAT	GTTGGCCGTC	GACGCGGTGT	TGGCGCCCGC	11100
	GTCGGGCGCC	GCGGAGCACA	GCAGCGCCGT	CGTGCCTCG	GCCATGTTGT	GCGCCAGCAC	11160
	CTGCAGCGTG	AGCATGGCGG	GCCCCGTCAAC	AACAACGCGC	CCGTTGTGGA	ACATGGCGTT	11220
	GACCGTGTTC	GCCACCAAAT	TGGCGGGATG	CAGCGGGTGG	GCGGGGTCCG	TCACGGGATC	11280
	GCTCGGGCAC	TCCTCGCCGG	GGGCGATCTC	CGGGACCACC	ATGTTCTGCA	GCGTGGCGTA	11340
10	CACGCGGTTC	AAGCGGACCC	CCGCGGTGCA	GCAGCGCCCC	CGCGAGAAGG	CCGGCACCAG	11400
	CACGTAATAG	TAGATTTTGT	GGTGGACGGT	CCAGTCGGCC	GGCCGGTGGC	GCCGGTCGTC	11460
	GGCGGCGTCG	GCCGCGCGGG	CCTGGGTGTT	GTGCAGCAGC	CGGCCGTCGT	TGCGGTTAAA	11520
	GTCGGCCGTC	GCCACGTTGC	ACGCCGCCGC	GTAGACGGGC	TCGTGTCCCC	CCGCGTCAAT	11580
	CCGGCAGTCT	CGGTGGCGGT	CCAGGGCCGC	GTGTCGCATA	AGGCCGTCCG	AGTCCCACAC	11640
15	GAGGGGCGGC	AGCAGCGCCG	GGTCGCGCAT	CAGGTGATTC	AGCTCGGCCT	GAGCCTGCCC	11700
	GCCCAGCTCC	GGCCCCGGCA	GGGTAAAGTC	GTCCACCAGC	TGGGCCAGGG	CCTCGACGTG	11760
	GGCCACCAGG	TCCCAGATCA	CGGCCATGCA	CTCCTCGGGG	AGGTCGCCCC	CGAGGTAGGT	11820
	CACGATGTAC	GAGACCAGCG	AGTAGTCGTT	CACGAACGCC	GCGCATCGCG	TGTTGTTCCA	11880
	GTAGCTGGTG	ATGCACTGAG	TCACGAGCCG	CGCCAGGGCG	CAGAACACGT	GCTCGCTGCC	11940
20	GTGAATCGCG	GCTTGCAGCA	GGTAAACAC	CGCCGGGTAG	CTGCGGTCCCT	CGAACGCCCC	12000
	GCGGACGGCG	GCTATGGTAG	CCGGCGCCAT	GGCGTGGCGG	CCAACGCCGA	GCTCCAGGCC	12060
	CCGGGCGTCA	CGAAACGCCA	CCGGACACAG	CGCCAGGGGC	AGGTGCCCCG	TGACCACGCG	12120
	CCAGGTGGCC	TGGATCGCCC	CCGGACCGGC	CGGGGGGACT	TCGCCGCCGG	GAAGCTCGAC	12180
	GTCGGCCACG	CCCGGAAGA	AGTCGAACGC	GGGTGTCAGC	TCCAGAGCCA	GGTTGGCGTT	12240
25	GTCGGGCTGC	ATGAACTGCT	CCGCGGTCAT	CTGGCACTCG	GCGACCCACC	GGACCCGGCC	12300
	GTGGGCGAGG	CGCTGCCGCC	AGGCGTTCAG	AAAACGCTGC	TGCATGTCCG	CGCCGGGGCC	12360
	GGCCGGGGCC	GCGACGTACG	CCCCGTACGG	ATTGCGGGCC	TCGACGGGGT	CGTGGTTTAC	12420
	GCCCCCGACG	GCCGCGTCGA	TGTTTCATGAG	CGAAGGATGA	CACACGGTCC	CGACCCGGTT	12480
	CTCCATGGAC	AGCCGCAGAA	CCTGGTGGTC	CTTTCCCCAA	AAAAACAGCT	GCCGGGGAGG	12540
30	GAACGCGCGG	GGCTCCGGGT	GGCCGGGGGC	GGGCACCAGG	TCCCCGGCGT	GCGCGGCGAA	12600
	GCGCTCCATG	GCCGGGTTGA	ACAGCCCCAG	GGGCAGGACG	AACGTCAGGT	CCATGGCGCC	12660
	CACCAGGGGG	TAGGGCACGT	TGGTGGCGGC	GTAGATGCGC	TTCTCCAGGG	CCTCCAGGAA	12720
	GACCAGCCTG	TCGCCTATGG	CCACCAGATC	CGCGCGCACG	CGCGTTGTCT	GGGGGGCGCT	12780
	TTCGAGTTCA	TCCAGCGTCT	CCCGGTTCCG	CTCGAGTTGC	TCCTCCTGCA	TCTCCAGCAG	12840
35	GTGGCGGCCC	ACGTCGTCCA	GGCTCCGCAC	GGCCTTGCCC	ATCACCAGCG	CCGTGACGAG	12900
	GTTGGCCCCG	TTCAAGACCA	TCTCGCCGTA	GGTCAACGGC	ACGTCGGCCT	CGGTGTCCCT	12960
	CACCTTCAGG	AAGGACTGCA	GGAGGCGCTG	TTTGATGGCG	GCGGTGGTGA	CCAGCACCCC	13020
	GTCGACCGGC	CGCCCGCGCG	TGTCGGCGTG	CGTCAGGCGG	GGCACGGCCA	CGGACGGCTG	13080
	CGTCGCGCTG	GTCAGGTCCA	CGAGCCAGGC	CTCGATGGCC	TCGCGGCGAT	GGCCCGCCTT	13140
40	GCCCAGGAAG	AAGCTCGTGT	CGAAAAGCT	CCGCTTCAGC	TCGGCGACCA	GGGTCGCCCC	13200
	GGCGACCCTG	GTCGCCAGGC	GCCCGTTGTC	GAGATATCGT	TGCATGGGCA	ACAGCAGGGC	13260
	CAGGGGAGGC	GCCTTCTCCA	ACAGCACGTG	CAGCATCTGG	TCGGCCGTGC	CGCGCTCAAA	13320
	CGCCCCCAGG	ACGGCCTGGA	CGTTGCGCGC	GAGCTGCTGG	ATGGCGCGCA	GCTGGCGATG	13380

	CAGGCTAATG	CCCGTCCCGT	CCAGGGCCTC	CCCCGTGAGC	AGGGCAATGG	CCTCGGTGGC	13440
	CAGGCTGAAG	GCGGCGTTCA	GGGCCCCGGC	GTCGATGACC	TTCGTCATGT	AATTATGCAC	13500
	GGGCTGCTCG	ACGGGGTGCG	GGCCGTCGCG	GGCGATGAGG	GGCTGGTGGA	CCTCGAACTG	13560
	CACACGCCCT	TCGTTTCATGT	AAGCCAGCTC	CGGGAACCTG	GTGCACACGC	ACGCCACGGA	13620
5	CAGGCCGAGC	TCCAGAAAGC	GCACGAGCGA	CAGGGTGTGT	CAGTAGGACC	CCAGCAGGGC	13680
	GTCAAACCTCT	ACGTCATACA	GGCTGTTTTT	GTCGGAGCGC	ACGCGGGCGA	AAAAATCAAA	13740
	GAGTCTGCGG	TGGGACGCCA	CCTCGATCGT	ACTCAGGATG	GAGCCGGTGG	GCACCATGGC	13800
	CGCGGCGTAC	CGGTAACCCG	GGGGGTGCGC	GGCAGGAGCG	GCCATTGGGT	TCCTTGGGGG	13860
	ATTCGAGGCG	TCCATCAAGC	CAAGCTCGGG	AAGGCCAAGC	CCCTCCCACA	CAACGCCTCA	13920
10	CCGCCGGCGG	ACGCGACTAA	CAACCCACGG	GCCGCCAAAA	ACCCCAAGGG	GCAACCCGAC	13980
	CAACAACAGG	CGAGGGGAGG	AAAGGCGTAA	AGGGGGCGTT	GGGAGGCAAA	AAGAAAGAAA	14040
	ACACCCAGAC	GTAGGCCCGA	GGACCGGCCG	GCGTCCTCTG	TCCCCGAGCA	CCCACTGTGC	14100
	CCAACAGGCA	CGGGGGCGAG	CTGCCCTGTC	CTTATATACC	CCCCCGCCAC	ACCCCCGTTA	14160
	GAACGCGACG	GGTGCCCTTCA	AGATGGCCCT	GGTCCAAAAG	CGTGCTAGAA	AAAAGTTGGT	14220
15	AAAGGCGGCA	AAGCAGTCCG	CCGCCGCCAC	CCACATGGCG	GCGCCGGCCG	CGCAGGCGAT	14280
	TCCCAGAGAA	CGGGCGCGGA	GGGGATCCGT	GCGGGGCAGC	AGCTGGCTGG	CGGTGATCCA	14340
	ATGGAAAAGC	CCGTCGGGAC	TGAACGTCTC	ATGGCGGGCC	GCCACCAGGG	CGCACAGGGC	14400
	CGCGCCGCCC	ATGATCACGC	ACAACCCCCA	AAACACGGGT	GGCGACAACG	GCAGGCGATC	14460
	CCGTTTGATG	TTCACGTACA	GGAGGAGCGC	CCGTGCCAGC	CACGTGACAT	AGTAGGCGAG	14520
20	GACGGCGGCT	ATAATACATG	CCGGCGCCAC	CGCCCGTCCG	GTCCACCCGT	AATACATGCC	14580
	CGCGGCCACC	AGCTCCAGCG	GCTTGAGGAC	CAGGAACGAC	CAAGCAAACA	TCACCACCCG	14640
	CTTGAAAAAG	ACCGGCTGGG	TGTGGGGCGG	AAGACGCGAG	TAGGCCGAAC	TGACAAAAAA	14700
	ATCAGACGTG	CCGTACGAGG	ACAGCGAAAA	CTGTTTCATCG	AGCGGCAGTT	CGCCGTCTCT	14760
	CCCCCCACAC	GCGGCCTCGT	ATACCAGCTC	GCGATCCAAC	AAAGGAACAT	CATCCCGCAT	14820
25	TGTCATGGTC	GGTGCGGGGA	GCCGGCGAGG	CAGCGAAACC	GAAAGTAGTG	CTGGCGGCGC	14880
	GGGCCCCGGT	CCGGACCCAA	GCTTCAGGGA	TGGGGGGCGG	AGGCCAAAAT	CAAACAAGCA	14940
	CCGCGCGGGT	TCTACACACA	ACCCCCACCC	GGGTAGTATC	CGCGGATGCG	AGTGCTTGGC	15000
	GAAGTCACGT	CCCAGCAGGA	TATAAACCTC	GGCCGTGTTG	CCCGGAACCC	CCGAAATTCA	15060
	CACCCACGCC	CTGACGCCCA	AATCATGGGT	GGATGTGGTT	CGCGAGCCGC	ACATCCGTGC	15120
30	GTCCGCCCTC	CCCCGCGGGC	TGATGACGTG	GCGGTTAGTC	AGTGGGAAGG	CAGGGGGAAA	15180
	GATGGGTTGG	GGGAGGAAAC	GAAAAAAACA	CCCAGAGGGC	CACGTCGGGA	ATGCGCCCGG	15240
	AGTTGTCTTT	AAAAGGCCGG	CCGTGCGTGA	CGGAAGCCGT	CGTTTGCCCA	AGCACCGACG	15300
	CCGCGATCCA	CAGTGGGGGG	AGTTCCTCCG	TCCGGCCACA	ACCTTACGCG	CGGGCGGCAC	15360
	GCGCGAGAGC	AACCCACGGG	TCCCGTTTCG	GCCACCGCCA	GCCCTTGCTC	CCACCACCTT	15420
35	CCTCCCACCA	CCCCACTATT	CCCCCCCCC	CAAGTCCGCC	CCGTGGCTCG	CCGGCCATGG	15480
	AGCTCACCTA	TGCCACCACC	CTGCACCACC	GGGACGTTGT	GTTTTACGTC	ACGGCAGACA	15540
	GAAACCGCGC	CTACTTTGTG	TGCGGGGGGT	CCGTTTATTC	CGTAGGGCGG	CCTCGGGATT	15600
	CTCAGCCGGG	GGAAATTGCC	AAGTTTGGCC	TGGTGGTCCG	GGGGACAGGC	CCCAAAGACC	15660
	GCATGGTTCG	CAACTACGTA	CGAAGCGAGC	TCCGCCAGCG	CGGCCGTGCG	GAAGTGCGGC	15720
40	CCGTGGGGGA	GGACGAGGTG	TTCTTGAGCA	GCGTGTGTCT	GCTAAACCCG	AACGTGAGCT	15780
	CCGAGCGAGA	CGTGATTAAT	ACCAACGACG	TTGAAGTGCT	GGACGAATGC	CTGGCCGAAT	15840
	ACTGCACCTC	GCTGCAAACC	AGCCCGGGGG	TGCTGGTGAC	CGGGGTGCGC	GTGCGCGCGC	15900
	GAGACAGGGT	CATCGAGCTA	TTTGAGCACC	CGGCGATCGT	CAACATTTCC	TCGCGCTTCG	15960

	CGTACACCCC	CTCCCCCTAC	GTATTCGCCC	TGGCCCAGGC	GCACCTCCCC	CGGCTCCCCA	16020
	GCTCGCTGGA	GCCCCTGGTG	AGCGGCCTGT	TTGACGGCAT	TCCCGCCCCG	CGCCAGCCCC	16080
	TGGACGCCCC	CGACCGGGCG	ACGGATGTTG	TGATCACGGG	CACCCGCGCC	CCCAGACCGA	16140
	TGGCCGGGAC	CGGGGCCGGG	GGCGCGGGGG	CCAAGCGGGC	CACCGTCAGC	GAGTTCGTGC	16200
5	AAGTGAAGCA	CATCGACCGT	GTGTGTGTC	CGAGCGTCTC	TTCCGCCCCC	CCGCCGAGCG	16260
	CCCCCGACGC	GAGTCTGCCG	CCCCCGGGG	TCCAGGAGGC	CGCCCCGCGG	GGCCCCCCGC	16320
	TCAGGGAGCT	GTGGTGGGTG	TTCTACGCCG	GCGACCGGGC	GCTGGAGGAG	CCCCACGCCG	16380
	AGTCGGGATT	GACGCGCGAG	GAGGTCCGCG	CCGTGCATGG	GTTCCGGGAG	CAGGCGTGGA	16440
	AGCTGTTTGG	GTCGGTGGGG	GCTCCGCGGG	CGTTTCTCGG	GGCCGCGCTG	GCCCTGAGCC	16500
10	CGACCCAAAA	GCTCGCCGTC	TACTACTATC	TCATCCACCG	GGAGCGGCGC	ATGTCCCCCT	16560
	TCCCCGCGCT	CGTGCGGCTC	GTCGGTCGGT	ACATCCAGCG	CCACGGCCTG	TACGTTCCCG	16620
	CGCCCCGACG	ACCGACGTTG	GCCGATGCCA	TGAACGGGCT	GTTCCGCGAC	GCGCTGGCGG	16680
	CCGGGACCGT	GGCCGAGCAG	CTCCTCATGT	TCGACCTCCT	CCCGCCCAAG	GACGTGCCGG	16740
	TGGGGAGCGA	CGCGCGGGCC	GACAGCGCCG	CCCTGCTGCG	CTTTGTGGAC	TCGCAACGCC	16800
15	TGACCCCGGG	GGGGTCCGTC	TCGCCCCGAG	ACGTCATGTA	CCTCGGCGCG	TTCTTGGGCG	16860
	TGTTGTACGC	CGGCCACGGA	CGCCTGGCCG	CGGCCACGCA	TACCGCGCGC	CTGACGGGCG	16920
	TGACGTCCCT	GGTCTTGACC	GTGGGGGACG	TCGACCGGAT	GTCCGCGTTT	GACCGCGGGC	16980
	CGGCGGGGGC	GGCTGGCCGC	ACGCGAACCG	CCGGGTACCT	GGACGCGCTG	CTTACCGTTT	17040
	GCCTGGCTCG	CGCCCAGCAC	GGCCAGTCTG	TGTGAGATAT	CCCAATAAAG	TGCAGTCGTT	17100
20	TTCTAACCAC	CGGATGCCGT	TGTATGCCTA	TACGGGGGAC	TATGGGGGGG	GGGGGAAAGG	17160
	AAAGGAAACA	GGAATGGAGA	AGGGAAAGGA	ACAGAGGCGG	TAGCGGACGC	ACGGCGGACA	17220
	CAATAACAAA	CAGACCGCGG	ACACGGAGGG	AGTCGGTTGG	GTTGGGCGTG	GACGCCGCTG	17280
	CGTCCACACA	CCCGTTTATT	CGCGTCTCCA	CAAAAATGGG	ACGCACGTTC	GGACCACCTT	17340
	GAGGATGCCC	GCCAGGGCCG	CGGTGATCAT	AACGACCCCC	AGCGCGGACG	CGGCCAGAAA	17400
25	CCCGGGGGCG	ATGGTGGCGA	TGGGCGAGCT	GTCAAAGGCC	AGCAGATGAA	TCACAGTTCC	17460
	GTGGGGGAAC	AACAACAGGG	CCACGGACGG	CACGTCGCTG	GAAAACACGT	TCGGGGTGCC	17520
	CGCCACCGGC	CCCTGGGCCA	GCTGCTGTTG	GGTGGCATCC	GTGTCCACCA	GCAGCACCGA	17580
	CATGACCTCC	CCGGCCGGGG	TGTAGCGCAG	AAACACGGCC	CCCACGAGGC	CGAGGTGCGG	17640
	CCGGTTTTTC	GTGCGCACCA	GCCGCTTCGG	CTCAATCTCC	CGCGCGTGCC	CTTCGCGAGT	17700
30	GGCGGTGAGA	TAGGTGATAA	ACAGCGGGCG	GCGGACGTCA	ACGCCCGTAA	GCTTGTATCC	17760
	GATCCCGCGG	GGCAAGGGGG	TGTGGGTGAC	GACGTAGCTG	GCGTTGTGGG	TGATGGGCAC	17820
	GAGGATCCCG	GGCTCCGCGT	TGTGCGACGG	GCCGCTACAC	TGGTGGGTGG	CTTCCGGGAC	17880
	GAAGGCGCGG	ATCAGGGCGT	TGTAGTGCGC	CCAGCGCGTG	AGAACGAGG	CCACGCCGCG	17940
	GGTCTGTTGT	GCCATGACGT	CCGCCGGGAT	GTCGGATCGG	GTGGCCATGG	CCAGCGCGTC	18000
35	CAGGATGAAC	CCGCCCTCGG	CGAGATCGAA	GCGCAGGGAA	GCTGCGCATG	GGGAAAAGTG	18060
	GTCCGGGAGC	CAGAAGAGGT	TTTTCTGGTG	GTCGGTCCTG	GCTAGCGCGG	CCCGGAGATC	18120
	GGCGTGGGTC	GCCGCGGCGA	CGTCGGACGT	ACACAGGGCC	GTGGTTATGA	GGAGGCCCGG	18180
	GCGGGCGCGT	TCCCGCTGCT	CGGCCGAGGG	CGCGCCCGCC	AGGAACGGCG	CCCGGAGGAC	18240
	GGCCGTGGCG	TAAACAGCG	CTCGCGGGAC	CATCGGGGCG	GTTAGCGCGC	GGCCGCCGAG	18300
40	AAACTCGGCG	TACAGGGCGT	CGATCAGGCG	GGCCGCGCTC	GGGGCCACCG	CGCCATAGGC	18360
	CGCGGGGCTG	TCCAACACGA	ACGCCAGCTG	ATAGCCCAGC	GCGTGCGCCG	CCAGGCTCTG	18420
	CTCTCGCTCG	AGGATCGCGG	CCACCAGATG	CCCGAGGCGC	GCCTCCAGCC	GCAGGCGGGC	18480
	CGCCGGGTCC	AACACGGACA	CGTTCAGGAA	CACCGAGTCG	GCCGCGCAGC	CCGCTGCTCC	18540



	CCGGGCGGCC	AGGCCGGCCA	GCACGCGCGA	GTGGGCCAAA	AAGCCCAGCA	GGTCGGAGAG	18600
	GCGAATCGCG	TCGTGGGCGT	GGGCCGCGTT	GACGAACGCA	AACCCCGACG	AGGCGAGCAG	18660
	CCCCGCGAGG	CGCCAGAACA	GGGACGGACG	CGCGTCCGTG	CCGGAGCCCG	GGTCCTCCCC	18720
	CAAAAACTCC	GCATAGGCC	GCGACATATA	CTGGGCGTAG	TTCGTGCTCT	CCTCGGGGTA	18780
5	GCCGGCCACC	CGCCGGAGGG	CGTCCAGCGC	CGAGCCGTTG	TCGGCGGGCG	TCGGGGCCCC	18840
	CAGGACAAAG	ACGCGATACC	TGGGGCCGGC	CGGAGGCCCG	GGGAGCACCG	CGGGGGCGTT	18900
	TTCTGCGGTC	GGATTTCCGA	CCCGAGCGAG	GGTCTTGTC	GCAGGCACCA	CTATGATCTC	18960
	GGCCGGAGGG	CTGTCCCGCA	TCGATATCAC	AACCCCCATG	AAGCCCTTCC	CGTATCGCGC	19020
	GCGCACAAGC	GCGGCGTCGC	ACCCGAACGC	CAGCCCGCCC	GTCGTCCAAA	CGCCACGGG	19080
10	CCACTTCAAG	GCCGACGGGG	AGAGGTACAC	TTACCGACCC	GGAGTCCGTA	GCAGGCCCTT	19140
	GGCGGCCAGC	CAGGTCACGG	ATGCGTTGTG	CAGATGCGCG	ATGCTCAGGT	TCGTGTCGG	19200
	ATGCCTCGGT	GTCCCCGCGG	GCGGCCCCGG	GGGCGGCGCG	TTGCGTCGGC	CGTCCGGGTG	19260
	CCTCTCGGTC	GCCCCGTCTG	CTCCCCGCGG	GAACGTAAGC	CCCTCGCGGT	CCGGCGCGGC	19320
	CGCGAATGTT	ACCCAGGCC	GGGACCGCAA	CAGCGCGGAG	GCGCCGGGGT	TGTGCGACAG	19380
15	TCCCTTGAGC	TGGGTCACCT	CGGCGGGGGG	ACGGGACGTG	GGCCCCGCCT	CGGGGAGCTC	19440
	GGGAGGCTC	GCGTTCCGAG	GCCGGCCGAG	CAGATAGGTC	TTTGGGATGT	AAAGCAGCTG	19500
	CCCGGGGTCC	CGAGGAAACT	CGGCCGTGGT	GACCAACACA	AAACAAAAGC	GCTCGGCGTA	19560
	CCACCGAAGC	ATGGGCACGG	ATGCCGTAGT	CAGGTTGAGT	TCGCCCCGGG	GCGCCAAGCG	19620
	TCCGCGCTGG	GGGTGCTGG	CGTCGGGGGT	GTTGGGCAAC	CACAGACGCC	CGGTGTTTGT	19680
20	GTCGCGCCAG	TACGTGCGGG	CCAACCCAG	ACCGTGCAAA	AACCACGGGT	CGATTGTGTC	19740
	CGTCCAGTAC	GTGTCATGGC	CCCCGGCAAC	GCCCACCAGG	ACCCCATCA	CCACCCACAG	19800
	ACCGGGGCCC	ATGGTCGTCC	GTCCCGGCTG	CCAGTCCGCA	GATGGGGGGG	TGTCCGTACC	19860
	CACGGCCCAA	AGAGGTCCG	CACCTCGGAG	GCTATCGGAG	GCCCTTTGTT	GCCGTAAGCG	19920
	CGGGCCAAAG	GATGGGGTGG	GGTGAGGGTA	AAAGCACAAA	GGGAGTACCA	GACCGAAAAC	19980
25	AAGGACGGAT	CGGCCCGCTC	CGTTTTCGCG	TGGGGTGCTG	ATACGGTGCC	AGCCCTGGCC	20040
	CCGAACCCCC	GCGCTTATGG	ACACACCACA	CGACAACAAT	GCCTTTTATT	CTGTCTCTTT	20100
	ATTGCCGTCA	TCGCCGGGAG	GCCTTCCGTT	CGGGCTTCCG	TGTTTGAAC	AACTCCCCC	20160
	CACCTCGCGG	GCAAACGTGC	GCGCCAGGTC	GCGTATCTCG	GCGATGGACC	CGGCGGTTGT	20220
	GACGCGGGTT	GGGATCATCC	CGGCGGTGAG	GCGCAACAGG	GCGTCTCGAC	ACCCGACGGG	20280
30	CGACTGATCG	TAATCCAGGA	CAAATAGATG	CATCGGAAGG	AGGCGGTCGG	CCAAGACGTC	20340
	CAAGACCCAG	GCAAAAATGT	GGTACAAGTC	CCCGTTGGGG	GCCAGCAGCT	CGGGAACGCG	20400
	GAACAGGGCA	AACAGCGTGT	CCTCGATGCG	GGGAGAGAC	CCCGCGCCGT	CCTCGGGGTC	20460
	GGGGCGCGGG	GTCGCCGCGG	CGACCCCGCT	CAGCCGGCCC	CAGTCTCCC	GCCACCTCCC	20520
	GCCGCGCTGC	AGGTACCGCA	CCGTGTTGGC	GAGTAGATCG	TAGACACGGC	GAATGGCGGA	20580
35	CAGCATGGCC	AGGTCAAGCC	GCTCGCCCGG	GCGTTGGCGT	CTGGCCAGGC	GGTCGGCGTG	20640
	TTGCGCCCTCC	GGAAGGACAC	CCAGGACCAG	GTTCTGCGCG	GGCGCGGTCG	GGGGCATGAG	20700
	GGCCACGAAC	GCCAACACGG	CCTGGGGGGT	CATGCTTCCC	ATGAGGTACC	GCGCGGCCGG	20760
	GTAGCACAGC	AGGGAGGCGA	TAGGGTGCCG	GTCGAAAACA	AGGGTGAGGG	CCGGGGGCGG	20820
	GGCTTGCGGG	CCCACAGCCT	CCCCCCCAT	ATGAGGAGCC	AAAACGGCGT	CCGTCGCCGC	20880
40	ATAAGGCGTG	CTCATGTGTA	TCTGGGCGCT	GGTCATTACC	ACCGCCGCCT	CCCCGGCCGA	20940
	TATCTCGCCG	CGGTCCAGAC	GGTGCTGCGT	GTTGTAGATG	TTCGTCAGGG	TCTCGGAGGC	21000
	CCCCAGCACC	TGCCAGTAAG	TCATCGGCTC	GGGACGCTAG	ACGATATTGT	CGCGCGGCC	21060
	CAGGGCCTCC	ATCAGCTGCG	CGGAGGTGGT	GGTCTTCCCC	ACCCCGTGGG	GTCCGTCTAT	21120

	ATAAACCCGC	AGCAGCGTGG	GCAGCTCCGG	ATCCCCGCGG	GCTTCGGAGG	CCCCCTGGCG	21180
	ATGGCTAGGA	CGGGACGCCG	CGCGGCCGTC	GGTAGGCCCG	CTCGCACGAG	CAGCCTGACC	21240
	GAACGCAGGC	GCGTGCTGTT	GGCCGGCGTG	AGAAGCCATA	CCCGCTTCTA	CAAGGCGTTC	21300
	CCCCGAGAGG	TGCGGGAGTT	CAACGCCACC	AGGATTGTG	GAACGCTGCT	GACGCTGATG	21360
5	AGCGGGTTCG	TGCAGGGTCG	CTCGCTGTTC	GAGGCCACGC	GCGTCACCTT	AATATGCGAA	21420
	GTGGACCTCG	GGCCGCGCCG	CCCAGACTGC	ATCTGCGTGT	TCGAATTCGC	CAATGACAAA	21480
	ACGTTGGGAG	GTGTGTGCGT	CATCCTGGAG	CTAAAGACAT	GCAAATCGAT	TTCTTCCGGG	21540
	GACACGGCCA	GCAAACGCGA	ACAGCGGACC	ACGGGCATGA	AGCAGCTGCG	CCACTCCCTG	21600
	AAGCTGCTGC	AGTCGCTCGC	GCCTCCGGGG	GACAAGGTCG	TCTACCTGTG	TCCTATTTTG	21660
10	GTGTTTGTG	CGCAGCGTAC	GCTGCGCGTC	AGCCGCGTGA	CCCGGCTCGT	CCCGCAAAAG	21720
	ATCTCCGGCA	ACATCACC GC	GGCCGTGCGG	ATGCTCCAAA	GCCTGTCCAC	GTATGCCGTG	21780
	CCGCCGGAAC	CGCAGACCCG	GCGGTGCGGG	CGCCGGGTTG	CCGCGACCGC	CAGACCGCAA	21840
	AGGCCCCCCT	CCCCGACACG	TGACCCGGAA	GGCACGGCGG	GTCATCCGGC	CCCACCAGAG	21900
	AGCGACCCCC	CTCCCCAGG	GGTTGTAGGC	GTCGCTGCGG	AGGGTGGGGG	TGTGCTTCAG	21960
15	AAAATCGCGG	CGCTTTTTCG	CGTGCCGGTG	GCCGCCAAGA	GCAGACCCCG	GACCAAAACC	22020
	GAGTGAGGTT	CTGTGTGTTG	TTTTTTTTC	TCGTTTTGTT	TTCTCTTCTT	TCCCCCCCCC	22080
	CTCCCCCGCT	TCTGGCCAAG	CATCCTCACC	TGCTTAAGCG	GAACCCGCGG	GCGCGCGGGG	22140
	ACTCATTTGT	CGCCGGCGAC	ACCCACCCGA	CAACAGCCCC	TGGGTGTAGA	CCGCTGTCGC	22200
	CCCCGTCTGT	CGCTCTCCC	TTTTTTTCCC	CCCTCAAAAA	ACGTGGTGTT	GGGCGCCGGC	22260
20	CAATTCCTCC	CGGAGCGCCG	TCGTGCGCCG	CCCGCCGCC	TCGAACATGG	ACCCGTACTA	22320
	CCCTTTCGAC	GCGCTGGACG	TTTGGGAACA	CAGGCGCTTC	ATCGTCGCCG	ACTCCAGGAG	22380
	CTTCATCACC	CCCGAGTTCC	CCCGGGACTT	CTGGATGTTG	CCCGTGTTCA	ACATCCCCCG	22440
	GGAGACGGCG	GCGGAGCGGG	CGGCAGTGAT	GCAGGCCAG	CGCACCGCGG	CCGCGGCGGC	22500
	CCTGGAGAAC	GCCGCCCTCC	AGGCCGCCGA	GCTGCCCGTC	GACATCGAGC	GCCGGATACG	22560
25	CCCGATCGAG	CAGCAGGTGC	ATCACATCGC	CGACGCCCTG	GAGGCGCTGG	AGACCGCGGC	22620
	GGCCGCGGCC	GAAGAGGCGG	ATGCCGCGCG	GGACGCCGAG	GCGAGGGGGG	AGGGCGCTGC	22680
	GGACGGGGCA	GCGCCGTCGC	CCACCGCGGG	CCCGCCGCC	GCGGAGATGG	AGGTTTCAGAT	22740
	CGTACGCAAC	GACCCGCCGC	TACGATACGA	TACCAACCTC	CCCGTGGATC	TGCTACACAT	22800
	GGTGACGCG	GGCCGCGGGG	CCGCGGGTTC	GTCGGGAGTC	GTCTTTGGTA	CCTGGTACCG	22860
30	CACGATCCAG	GAACGCACCA	TCGCGGACTT	CCCCCTGACC	ACCCGCAGCG	CCGACTTTCG	22920
	AGACGGGCGC	ATGTCCAAAA	CCTTCATGAC	CGCGCTGGTC	CTGTCTCTGC	AGTCGTGCGG	22980
	CCGGCTGTAC	GTGGGCCAGC	GCCACTATTC	CGCTTTCGAG	TGCGCCGTGC	TGTGTCTGTA	23040
	TCTGCTGTAC	CGAACCACCC	ACGAGTCCTC	CCCGATCGC	GATCGCGCTC	CCGTTGCGTT	23100
	CGGGGACCTG	CTGGCCCGCC	TGCCGCGCTA	CCTGGCGCGT	CTGGCCGCGG	TAATCGGCGA	23160
35	CGAGAGCGGA	CGCCCGCAGT	ACCGCTACCG	CGACGACAAG	CTGCCCAAAG	CGCAGTTCGC	23220
	GGCGGCCGGC	GGCCGCTACG	AGCACGGGGC	CCTGGCCACC	CACGTCGTGA	TCGCCACGTT	23280
	GGTGCGCCAC	GGGTGCTAC	CGGCGGCCCC	GGGCGACGTT	CCCGAGACA	CCAGCACCCG	23340
	CGTGAACCCC	GACGACGTGG	CCCACCGCGA	CGACGTCAAC	CGCGCCGCCG	CCGCGTTTTT	23400
	GGCACGCGGC	CACAACCTCT	TCCTGTGGGA	GGACCAGACG	CTGCTGCGGG	CGACCGCCAA	23460
40	CACCATTACG	GCCCTGGCCG	TGCTTCGGCG	GCTCCTCGCG	AACGGCAACG	TGTACGCGGA	23520
	CCGCTCGAC	AACCGCCTGC	AGCTGGGCAT	GCTGATCCCG	GGAGCCGTCC	CGGCGGAGGC	23580
	CATCGCTCGG	GGGGCGTCCG	GATTGGACTC	GGGCGCCATA	AAAAGCGGCG	ACAACAACCT	23640
	GGAGGCGCTG	TGCGTTAACT	ATGTACTTCC	GCTGTATCAG	GCAGACCCCA	CGGTCGAGCT	23700

	GACCCAGTTG	TTTCCGGGGG	CTGGCCGCCC	TGTGCCTGGA	CGCCCAGGCG	GGGCGGCCAC	23760
	TGGCGTTGAC	GAGGCGCGTG	GTGGATATGT	TGTCGGGCGC	CCGCCAGGCG	GCGCTCGTGC	23820
	GCCTCACCGC	GCTGGAGCTC	ATCAACCGCA	CCCGCACAAA	CACCACCCCT	GTGGGGGAGA	23880
	TTATTAACGC	CCACGATGCC	TTGGGGATAC	AATACGAACA	GGGCCTGGGG	CTGCTCGCCC	23940
5	AGCAGGCACG	CATCGGCTTG	GCGTCGAACG	CCAAGCGATT	CGCCACGTTC	AACGTGGGCA	24000
	GCGACTACGA	CCTGTGTGAC	TTTTTGTGTC	TCGGGTTCAT	TCCCCAGTAC	CTGTCCGTGG	24060
	CCTAGGGAAG	GGTGGGGGTG	GTGGTGGTGG	GGTGTPTTTC	TGCTGTGTGT	GTTTCTGGTC	24120
	GCGCTGGTCA	CAAAAGGCAC	GGCGCCCCGA	AACGCGGGCT	TTAGTCCCGG	CCCGGACGTC	24180
	GGCGGACACA	CAACAACGGC	GGGCCCCGTG	GGTGGGTAAG	TTGGTTCGGG	GGCATCGCTG	24240
10	TATTCCTTG	CCCGCTTCCA	CCCCCCTTC	CCGTTTTGTT	TGTTTTGTGCG	GGTGCCCATG	24300
	GCGTCGGCGG	AAATGCGCGA	GCGGTTGGAG	GCGCCTCTGC	CCGACCGGGC	GGTGCCCATC	24360
	TACGTGGCCG	GGTTTTTGGC	CCTGTACGAC	AGCGGGGACC	CGGGCGAGCT	GGCCCTGGAC	24420
	CCAGACACGG	TGCGTGCGGC	CCTGCCTCCG	GAGAACCCCC	TGCCGATCAA	CGTAGACCAC	24480
	CGCGCTCGGT	GCGAGGTGGG	CCGGGTGCTC	GCCGTGGTCA	ACGACCCTCG	GGGGCCGTTT	24540
15	TTTGTGGGGC	TGATCGCGTG	CGTGACAGTG	GAGCGCGTCC	TCGAGACGGC	CGCCAGCGCC	24600
	GCTATTTTTG	AGCGCCGCGG	ACCCGCGCTC	TCCCGGGAGG	AGCGTCTGCT	GTACCTGATC	24660
	ACCAACTACC	TGCCATCGGT	CTCGCTGTCC	ACAAAACGCC	GGGGGACGA	GGTTCCGCCC	24720
	GACCGCACCC	TGTTTGCGCA	CGTGGCCCTG	TGCGCCATCG	GGCGGCGCCT	TGGAACCATC	24780
	GTCACCTACG	ACACCAGCCT	AGACGCGGCC	ATCGCTCCGT	TTCGCCACCT	GGACCCGGCG	24840
20	ACGCGCGAGG	GGGTGCGACG	CGAGGCCGCC	GAGGCCGAGC	TCGCGCTGGC	CGGGCGCACC	24900
	TGGGCCCCCG	GCGTGGAGGC	GCTCACACAC	ACGCTGCTCT	CCACCGCCGT	CAACAACATG	24960
	ATGCTGCGTG	ACCGCTGGAG	CCTCGTGGCC	GAGCGGCGGC	GGCAGGCCGG	GATCGCCGGA	25020
	CACACGTACC	TTCAGGCGAG	CGAAAAATTT	AAAATATGGG	GGGCGGAGTC	TGCCCTTGCG	25080
	CCGGAGCGCG	GGTATAAAAC	CGGCGCCCCG	GGTGCCATGG	ACACATCCCC	CGCCGCGAGC	25140
25	GTTCCCGCGC	CGCAGGTCGC	CGTCCGTGCG	CGTCAAGTCG	CGTCGTCGTC	GTCTTCTTCT	25200
	TCTTCTTTTC	CGGCACCGGC	CGATATGAAC	CCCGTTTCGG	CATCGGGCGC	CCCGGCCCCCT	25260
	CCGCCGCCCC	GCGACGGGAG	TTATTTGTGG	ATCCCCGCCT	TTCATTACAA	TCAGCTCGTC	25320
	ACCGGGCAAT	CCGCGCCCCA	CCACCCGCGC	CTGACCGCGT	GCGGCCTGCC	GGCCGCGGGG	25380
	ACGGTGGCCT	ACGGACACCC	CGGCGCCGGC	CCGTCCCCGC	ACTACCCGCC	TCCTCCCGCC	25440
30	CACCCGTACC	CGGGGTATGC	TGTTTCGCGG	CCCCAGTCCC	CTGGAGGCCC	AGATCGCCGC	25500
	GCTGGTGGGG	GCCATCGCCG	CCGACCGCCA	GGCGGGTGGG	CTTCCGGCGG	CCGCCGGAGA	25560
	CCACGGGATC	CGGGGTTCGG	CGAACCGCCG	CCGACACGAG	GTGGAGCAGC	CGGAGTACGA	25620
	CTGCGGCCGT	GACGAGCCGG	ACCGGGACTT	CCCGTATTAC	CCGGGCGAGG	CCCGCCCCGA	25680
	GCCGCGCCCC	GTCGACTCCC	GGCGCGCCGC	GCGCCAGGCT	TCCGGGCCCC	ACGAAACCAT	25740
35	CACGGCGCTG	GTGGGGGCGG	TGACGTCCCT	GCAGCAGGAA	CTGGCGCACA	TGCGCGCGCG	25800
	TACCCACGCC	CCCTACGGGC	CGTATCCGCC	GGTGGGGCCC	TACCACCACC	CCCACGCAGA	25860
	CACGGAGACC	CCCCCCCAAC	CACCCCGCTA	CCCCGCCGAG	GCCGTCTATC	TGCCGCGGCC	25920
	GCACATCGCC	CCCCCGGGGC	CTCCTCTATC	CGGGGCGGTC	CCCCCACCCT	CGTATCCCCC	25980
	AGTTGCGGTT	ACCCCCGGTC	CCGCCCCCCC	GCTACATCAG	CCCTCCCCCG	CACACGCCCA	26040
40	CCCCCTCCG	CCGCCGCCGG	GACCCACGCC	TCCCCCGGCC	GCGAGCTTAC	CCCAACCCGA	26100
	GGCGCCCGGC	GCGGAGGCCG	GCGCCTTAGT	TAACGCCAGC	AGCGCGGCCC	ACGTGAACGT	26160
	GGACACGGCC	CGGGCCGCCG	ATTTGTTTGT	GTCACAGATG	ATGGGGTCCC	GCTAACTCGC	26220
	CTCCAGGATC	CGGACTTGGG	GGGGGTGTGT	GTTTTCATAT	ATTTTAAATA	AACAAACAAC	26280

CGGACAAAAG TATACCCACT TCGTGTGCTT GTGTTTTTGT TTGAGAGGGG GGGGGTGG 26339

(2) INFORMATION FOR SEQ ID NO:38:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 897 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

15 Val Ser Gly Arg Ala Gly Asp Pro Ala Gly Leu Pro Ala Pro Arg Gly  
     1                    5                    10                    15  
 Gly Pro Thr Trp Pro Met Pro Ser Gly Gly Pro Pro Pro Glu Val Lys  
                     20                    25                    30  
 Ala Gly Leu Arg Ala Asp Met Trp Gly Val Met Gly Gln Tyr Arg Glu  
 20                    35                    40                    45  
 Ala Xaa Glu His Gln Thr Pro Asp Thr Glu Thr Val Val Ala Gly Met  
     50                    55                    60  
 His Pro Ala Leu Val Val Val Leu Lys Thr Met Phe Xaa Asp Ala Pro  
     65                    70                    75                    80  
 25 Glu Thr Pro Val Leu Val Gln Phe Phe Ser Asp His Ala Pro Thr Ile  
                     85                    90                    95  
 Ala Lys Ala Val Ser Asn Ala Ile Asn Ala Gly Ser Ala Ala Val Ala  
                     100                    105                    110  
 Thr Asp Ala Ala Thr Val Asp Ala Ala Val Arg Ala His Gly Ala Asp  
 30                    115                    120                    125  
 Ala Val Ser Ala Leu Gly Ala Ala Ala Arg Asp Pro Asp Leu Ser Phe  
     130                    135                    140  
 Leu Ala Ala Asp Ser Ala Ala Gly Tyr Val Lys Ala Thr Arg Leu Ala  
     145                    150                    155                    160  
 35 Leu Glu Arg Ala Ile Asp Lys Leu Thr Thr Leu Gly Ser Ala Ala Ala  
                     165                    170                    175  
 Asp Leu Val Phe His Ala Arg Arg Ala Cys Ala Gln Pro Glu Gly Asp  
     180                    185                    190  
 His Ala Ala Leu Ile Asp Ala Ala Ala Arg Ala Thr Thr Ala Ala Arg  
 40                    195                    200                    205  
 Glu Ser Leu Ala Gly His Glu Ala Gly Phe Gly Gly Leu Leu His Ala  
     210                    215                    220  
 Glu Gly Thr Ala Gly Asp His Ser Pro Ser Gly Arg Ala Leu Gln Glu  
                     237

225                      230                      235                      240  
 Leu Gly Lys Val Ile Gly Ala Thr Arg Arg Arg Ala Glu Glu Leu Glu  
                                  245                      250                      255  
 Ala Ala Val Ala Asp Leu Thr Gly Lys Met Ala Ala Gln Arg Arg Ser  
 5                      260                      265                      270  
 Ser Trp Ala Ala Gly Val Glu Ala Ala Leu Asp Arg Val Glu Asn Arg  
                                  275                      280                      285  
 Ala Glu Phe Asp Val Val Glu Leu Arg Arg Leu Gln Ala Gly Thr His  
                                  290                      295                      300  
 10 Gly Tyr Asn Pro Arg Asp Phe Arg Lys Arg Ala Glu Gln Ala Ala Asn  
 305                      310                      315                      320  
 Ala Glu Ala Val Thr Leu Ala Leu Asp Thr Ala Phe Ala Phe Asn Pro  
                                  325                      330                      335  
 Tyr Thr Pro Glu Asn Gln Arg His Pro Met Leu Pro Pro Leu Ala Ala  
 15                      340                      345                      350  
 Ile His Arg Leu Gly Trp Ser Ala Ala Phe His Ala Ala Ala Glu Thr  
                                  355                      360                      365  
 Tyr Ala Asp Met Phe Arg Val Asp Ala Glu Pro Leu Ala Arg Leu Leu  
                                  370                      375                      380  
 20 Arg Ile Ala Glu Gly Leu Leu Glu Met Ala Gln Ala Gly Asp Gly Phe  
 385                      390                      395                      400  
 Ile Asp Tyr His Glu Ala Val Gly Arg Leu Ala Asp Asp Met Thr Ser  
                                  405                      410                      415  
 Val Pro Gly Leu Arg Arg Tyr Val Pro Phe Phe Gln His Gly Tyr Ala  
 25                      420                      425                      430  
 Asp Tyr Val Glu Leu Arg Asp Arg Leu Asp Ala Ile Arg Ala Asp Val  
                                  435                      440                      445  
 His Arg Ala Leu Gly Gly Val Pro Leu Asp Leu Ala Ala Ala Ala Glu  
                                  450                      455                      460  
 30 Gln Ile Ser Ala Ala Arg Asn Asp Pro Glu Ala Thr Ala Glu Leu Val  
 465                      470                      475                      480  
 Arg Thr Gly Val Thr Leu Pro Cys Pro Ser Glu Asp Ala Leu Val Ala  
                                  485                      490                      495  
 Cys Ala Ala Ala Leu Glu Arg Val Asp Gln Ser Pro Val Lys Asn Thr  
 35                      500                      505                      510  
 Ala Tyr Ala Glu Tyr Val Ala Phe Val Thr Arg Gln Asp Thr Ala Glu  
                                  515                      520                      525  
 Thr Lys Asp Ala Val Val Arg Ala Lys Gln Gln Arg Ala Glu Ala Thr  
                                  530                      535                      540  
 40 Glu Arg Val Met Ala Gly Leu Arg Glu Ala Ala Arg Glu Arg Arg Ala  
 545                      550                      555                      560  
 Gln Ile Glu Ala Glu Gly Leu Ala Asn Leu Lys Thr Met Leu Lys Val  
                                  565                      570                      575

Val Ala Val Pro Ala Thr Val Ala Lys Thr Leu Asp Gln Ala Arg Ser  
 580 585 590  
 Val Ala Glu Ile Ala Asp Gln Val Glu Val Leu Leu Asp Gln Thr Glu  
 595 600 605  
 5 Lys Thr Arg Glu Leu Asp Val Pro Ala Val Ile Trp Leu Glu His Ala  
 610 615 620  
 Gln Arg Thr Phe Glu Thr His Pro Leu Ser Ala Arg Asp Gly Pro Gly  
 625 630 635 640  
 Pro Leu Ala Arg His Ala Gly Arg Leu Gly Ala Leu Phe Asp Thr Arg  
 10 645 650 655  
 Arg Arg Val Asp Ala Leu Arg Arg Ser Leu Glu Glu Ala Glu Ala Glu  
 660 665 670  
 Trp Asp Glu Val Trp Gly Arg Phe Gly Arg Val Arg Gly Gly Ala Trp  
 675 680 685  
 15 Lys Ser Pro Glu Gly Phe Arg Ala Met His Glu Gln Leu Arg Ala Leu  
 690 695 700  
 Gln Asp Thr Thr Asn Thr Val Ser Gly Leu Arg Ala Gln Pro Ala Tyr  
 705 710 715 720  
 Glu Arg Leu Ser Ala Arg Tyr Gln Gly Val Leu Gly Ala Lys Gly Ala  
 20 725 730 735  
 Glu Arg Ala Glu Ala Val Glu Glu Leu Gly Ala Arg Val Thr Lys His  
 740 745 750  
 Thr Ala Leu Cys Ala Arg Leu Arg Asp Glu Val Val Arg Arg Val Pro  
 755 760 765  
 25 Trp Glu Met Asn Phe Asp Ala Leu Gly Arg Leu Leu Ala Glu Phe Asp  
 770 775 780  
 Ala Ala Ala Ala Asp Leu Ala Pro Trp Ala Val Glu Glu Phe Arg Gly  
 785 790 795 800  
 Ala Arg Glu Leu Ile Gln Tyr Arg Met Gly Ser Ala Tyr Ala Arg Ala  
 30 805 810 815  
 Gly Gly Lys Ala Leu Phe Leu Phe Phe Phe Phe Pro Pro Pro Leu Ser  
 820 825 830  
 Ser Phe Leu Pro His Phe His Phe Phe Ile His His His His Ser Phe  
 835 840 845  
 35 Thr Lys Phe Phe Thr Ser Ser Ser Leu His Ser Tyr His Leu Phe Pro  
 850 855 860  
 Ser Ser Ile Tyr Ser Ile Pro Ser Ile Ser Pro Leu Tyr Pro His Ser  
 865 870 875 880  
 Ser Leu Ser Phe Pro Ser Ser Gln Phe Leu His Ile Phe Leu Ser Leu  
 40 885 890 895  
 Pro

## (2) INFORMATION FOR SEQ ID NO:39:

## (i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 335 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Val Met Pro Val Ala Pro Pro Pro Arg Gly Ala Gly Gly Arg Ala Pro  
 1 5 10 15  
 15 Cys Pro Pro Ala Leu Gly Pro Glu Ala Ile His Ala Arg Leu Glu Asp  
 20 25 30  
 Val Arg Ile Gln Ala Arg Arg Ala Ile Glu Ser Ala Ile Lys Glu Tyr  
 35 40 45  
 Phe His Arg Gly Ala Val Tyr Ser Ala Lys Ala Leu Gln Ala Ser Asp  
 20 50 55 60  
 Ser His Asp Cys Arg Phe His Val Ala Ser Ala Ala Val Val Pro Met  
 65 70 75 80  
 Val Gln Leu Leu Glu Ser Leu Pro Ala Phe Asp Gln His Thr Arg Asp  
 85 90 95  
 25 Val Ala Gln Arg Ala Ala Leu Pro Pro Pro Pro Pro Leu Ala Thr Ser  
 100 105 110  
 Pro Gln Ala Ile Leu Leu Arg Asp Leu Leu Gln Arg Gly Gln Thr Leu  
 115 120 125  
 Asp Ala Pro Glu Asp Leu Ala Ala Trp Leu Ser Val Leu Thr Asp Ala  
 30 130 135 140  
 Ala Thr Gln Gly Leu Ile Glu Arg Lys Pro Leu Glu Glu Leu Ala Arg  
 145 150 155 160  
 Ser Ile His Gly Ile Asn Asp Gln Gln Ala Arg Arg Ser Ser Gly Leu  
 165 170 175  
 35 Ala Glu Leu Gln Arg Phe Asp Ala Leu Asp Ala Ala Gln Gln Leu Asp  
 180 185 190  
 Ser Asp Ala Ala Phe Val Pro Ala Thr Gly Pro Ala Pro Tyr Val Asp  
 195 200 205  
 Gly Gly Gly Leu Ser Pro Glu Ala Thr Arg Met Ala Glu Asp Ala Leu  
 40 210 215 220  
 Arg Gln Ala Arg Ala Met Glu Ala Ala Lys Met Thr Ala Glu Leu Ala  
 225 230 235 240  
 Pro Glu Ala Arg Ser Arg Leu Arg Glu Arg Ala His Ala Leu Glu Ala

```

                245                250                255
Met Leu Asn Asp Ala Arg Glu Arg Ala Lys Val Ala His Asp Ala Arg
                260                265                270
Glu Lys Phe Leu His Lys Leu Gln Gly Val Leu Arg Pro Leu Pro Asp
5      275                280                285
Phe Val Gly Leu Lys Ala Cys Pro Ala Val Leu Ala Thr Leu Arg Ala
      290                295                300
Ser Leu Pro Arg Gly Val Asp Arg Pro Gly Arg Cys Arg Pro Gly Ala
305                310                315                320
10    Pro Pro Arg Lys Ser Arg Arg Gly Cys Gly Arg Thr Cys Gly Gly
      325                330                335

```

## (2) INFORMATION FOR SEQ ID NO:40:

## 15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 800 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

```

25 Val Val Thr Gly Val Arg Asn Gln Phe Ala Thr Asp Leu Glu Pro Gly
   1          5          10          15
Gly Ser Val Ser Cys Met Arg Ser Ser Leu Ser Phe Leu Ser Leu Leu
   20          25          30
Phe Asp Val Gly Pro Arg Asp Val Leu Ser Ala Glu Ala Ile Glu Gly
30  35          40          45
Cys Leu Val Glu Gly Gly Glu Trp Thr Arg Ala Ala Ala Gly Ser Gly
   50          55          60
Pro Pro Arg Met Cys Ser Ile Ile Glu Leu Pro Asn Phe Leu Glu Tyr
65          70          75          80
35 Pro Ala Arg Gly Leu Arg Cys Val Phe Ser Arg Val Tyr Gly Glu Val
   85          90          95
Gly Phe Phe Gly Glu Pro Thr Ala Gly Leu Leu Glu Thr Gln Cys Pro
   100         105         110
Ala His Thr Phe Phe Ala Gly Pro Trp Ala Met Arg Pro Leu Ser Tyr
40  115         120         125
Thr Leu Leu Thr Ile Gly Pro Leu Gly Met Gly Arg Asp Gly Asp Thr
   130         135         140
Ala Tyr Leu Phe Asp Pro His Gly Leu Pro Ala Gly Thr Pro Ala Phe

```



145                      150                      155                      160  
 Ile Ala Lys Val Arg Ala Gly Asp Val Tyr Pro Tyr Leu Thr Tyr Tyr  
                                  165                      170                      175  
 Ala His Asp Arg Pro Lys Val Arg Trp Ala Gly Ala Met Val Phe Phe  
 5                      180                      185                      190  
 Val Pro Ser Gly Pro Gly Ala Val Ala Pro Ala Asp Leu Thr Ala Ala  
                                  195                      200                      205  
 Ala Leu His Leu Tyr Gly Ala Ser Glu Thr Tyr Leu Gln Asp Glu Pro  
                                  210                      215                      220  
 10 Phe Val Glu Arg Arg Val Ala Ile Thr His Pro Leu Arg Gly Glu Ile  
                                  225                      230                      235                      240  
 Gly Gly Leu Gly Ala Leu Phe Val Gly Val Val Pro Arg Gly Asp Gly  
                                  245                      250                      255  
 Glu Gly Ser Gly Pro Val Val Pro Ala Leu Pro Ala Pro Thr His Val  
 15                      260                      265                      270  
 Gln Thr Pro Arg Ala Asp Arg Pro Pro Glu Ala Pro Arg Gly Ala Ser  
                                  275                      280                      285  
 Gly Pro Pro Asn Thr Pro Gln Ala Gly His Pro Asn Arg Pro Pro Asp  
                                  290                      295                      300  
 20 Asp Val Trp Ala Ala Ala Leu Glu Gly Thr Pro Pro Ala Lys Pro Ser  
                                  305                      310                      315                      320  
 Ala Pro Asp Ala Ala Ala Ser Gly Pro Pro His Ala Ala Pro Pro Pro  
                                  325                      330                      335  
 Gln Thr Pro Ala Gly Asp Ala Ala Glu Glu Ala Glu Asp Leu Arg Val  
 25                      340                      345                      350  
 Leu Glu Val Gly Ala Val Pro Val Gly Cys His Arg Ala Arg Tyr Ser  
                                  355                      360                      365  
 Thr Gly Leu Pro Lys Arg Arg Arg Pro Thr Trp Thr Pro Pro Ser Ser  
                                  370                      375                      380  
 30 Val Glu Asp Leu Thr Ser Gly Glu Arg Pro Ala Pro Lys Ala Pro Pro  
                                  385                      390                      395                      400  
 Ala Lys Ala Lys Lys Lys Ser Ala Pro Lys Lys Lys Ala Pro Val Ala  
                                  405                      410                      415  
 Ala Glu Val Pro Ala Ser Ser Pro Thr Pro Ile Ala Ala Thr Val Pro  
 35                      420                      425                      430  
 Pro Ala Pro Asp Thr Pro Pro Gln Ser Gly Gln Gly Gly Gly Asp Asp  
                                  435                      440                      445  
 Gly Pro Asp Ser Ser Pro Ser Val Leu Glu Thr Leu Gly Ala Arg Arg  
                                  450                      455                      460  
 40 Pro Pro Glu Pro Pro Gly Ala Asp Leu Ala Gln Leu Phe Glu Val His  
                                  465                      470                      475                      480  
 Pro Asn Val Ala Ala Thr Ala Val Arg Leu Ala Ala Arg Asp Ala Ala  
                                  485                      490                      495

Arg Glu Val Ala Ala Cys Ser Gln Leu Thr Ile Asn Ala Leu Arg Ser  
 500 505 510  
 Pro Tyr Pro Ala His Pro Gly Leu Leu Glu Leu Cys Val Ile Phe Phe  
 515 520 525  
 5 Phe Glu Arg Val Leu Ala Phe Leu Ile Glu Asn Gly Ala Arg Thr His  
 530 535 540  
 Thr Gln Ala Gly Val Ala Gly Pro Ala Ala Ala Leu Leu Asp Phe Thr  
 545 550 555 560  
 Leu Arg Met Pro Pro Arg Lys Thr Ala Val Gly Asp Phe Leu Ala Ser  
 10 565 570 575  
 Thr Arg Met Ser Leu Ala Asp Val Ala Ala His Arg Pro Leu Ile Gln  
 580 585 590  
 His Val Leu Asp Lys Asn Ser Gln Ile Gly Arg Leu Ala Lys Leu Val  
 595 600 605  
 15 Leu Val Ala Arg Asp Phe Ile Arg Glu Thr Asp Ala Phe Tyr Gly Asp  
 610 615 620  
 Leu Ala Asp Leu Asp Leu Gln Leu Arg Ala Ala Pro Pro Ala Asn Leu  
 625 630 635 640  
 Tyr Ala Arg Leu Gly Lys Trp Leu Leu Glu Arg Ser Arg Ala His Pro  
 20 645 650 655  
 Asn Thr Leu Phe Ala Pro Ala Thr Pro Thr His Pro Glu Pro Leu Leu  
 660 665 670  
 His Arg Ile Gln Ala His Phe Arg Lys Lys Met Arg Val Glu Ala Glu  
 675 680 685  
 25 Ala Arg Glu Met Arg Glu Ala Leu Tyr Arg Val Tyr Ser Val Ser Gln  
 690 695 700  
 Arg Ala Gly Pro Pro Asp Arg Asp Ala Arg Cys Pro Pro Pro Gly  
 705 710 715 720  
 Arg Arg Arg Gln Gly Pro Val Pro Ala Arg Pro Gly Pro Arg Gly His  
 30 725 730 735  
 Pro Cys Ala Ala Gly Gly Arg Ala Asp Pro Gly Pro Pro Gly Asp Arg  
 740 745 750  
 Lys Arg Asp Gln Gly Val Leu Pro Pro Gly Ser Arg Ile Gln Arg Glu  
 755 760 765  
 35 Gly Pro Ala Gly Gln Arg Gln Pro Arg Leu Ser Val Ser Arg Gly Leu  
 770 775 780  
 Gly Arg Gly Arg Ala His Gly Pro Val Ala Gly Ile Ala Thr Gly Leu  
 785 790 795 800

40

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 158 amino acids

(B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

```

10 Met Asn Ala His Phe Ala Asn Glu Val Gln Tyr Asp Leu Thr Arg Asp
    1           5           10           15
    Pro Ser Ser Pro Ala Ser Leu Ile His Val Ile Ile Ser Ser Glu Cys
        20           25           30
    Leu Ala Ala Ala Gly Val Pro Leu Ser Ala Leu Val Arg Gly Arg Pro
        35           40           45
15 Asp Gly Gly Ala Ala Ala Asn Phe Arg Val Glu Thr Gln Trp His Ala
    50           55           60
    Pro Gly Asp Cys Thr Pro Trp Arg Ser Ala Phe Ala Ala Tyr Val Pro
    65           70           75           80
    Ala Asp Ala Val Gly Ala Ile Leu Ala Pro Val Ile Pro Ala His Pro
20           85           90           95
    Asp Leu Leu Pro Arg Val Pro Ser Ala Gly Gly Leu Phe Val Ser Leu
        100           105           110
    Pro Val Ala Cys Asp Ala Gln Gly Val Tyr Asp Pro Tyr Thr Val Ala
        115           120           125
25 Ala Leu Arg Leu Ala Trp Gly Pro Trp Ala Thr Cys Ala Arg Val Leu
    130           135           140
    Leu Phe Ser Tyr Asp Glu Leu Thr Arg Tyr Arg Val Cys Gly
    145           150           155

```

30 (2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 423 amino acids  
 (B) TYPE: amino acid  
 35 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

```

Val Pro Glu Gly Ala Trp Val Gly Gly Ala Cys Ala Arg Pro Arg Gly
1           5           10           15

```

Pro Arg Ala His Val Arg Leu Tyr Ala Val Cys Phe Val Cys Pro Gln  
 20 25 30  
 Gly Ile Arg Gly Gln Asp Phe Asn Leu Leu Phe Val Asp Glu Ala Asn  
 35 40 45  
 5 Phe Ile Arg Pro Asp Ala Val Gln Thr Ile Met Gly Phe Leu Asn Gln  
 50 55 60  
 Ala Asn Cys Lys Ile Ile Phe Val Ser Ser Thr Asn Thr Gly Lys Ala  
 65 70 75 80  
 Ser Thr Ser Phe Leu Tyr Asn Leu Arg Gly Ala Ala Asp Glu Leu Leu  
 85 90 95  
 10 Asn Val Val Thr Tyr Ile Cys Asp Asp His Met Pro Arg Val Val Thr  
 100 105 110  
 His Thr Asn Ala Thr Ala Cys Ser Cys Tyr Ile Leu Asn Lys Pro Val  
 115 120 125  
 15 Phe Ile Thr Met Asp Gly Ala Val Arg Arg Thr Ala Asp Leu Phe Leu  
 130 135 140  
 Pro Asp Ser Phe Met Gln Glu Ile Ile Gly Gly Gln Ala Arg Glu Thr  
 145 150 155 160  
 Gly Asp Asp Arg Pro Val Leu Thr Lys Ser Ala Gly Glu Arg Phe Leu  
 165 170 175  
 20 Leu Tyr Arg Pro Ser Thr Thr Thr Asn Ser Gly Leu Met Ala Pro Glu  
 180 185 190  
 Leu Tyr Val Tyr Val Asp Pro Ala Phe Thr Ala Asn Thr Arg Ala Ser  
 195 200 205  
 25 Gly Thr Gly Ile Ala Val Val Gly Arg Tyr Arg Asp Asp Phe Ile Ile  
 210 215 220  
 Phe Ala Leu Glu His Phe Phe Leu Arg Ala Leu Thr Gly Ser Ala Pro  
 225 230 235 240  
 Ala Asp Ile Ala Arg Cys Val Val His Ser Leu Ala Gln Val Leu Ala  
 245 250 255  
 30 Leu His Pro Gly Ala Phe Arg Ser Val Arg Val Ala Val Glu Gly Asn  
 260 265 270  
 Ser Ser Gln Asp Ser Ala Val Ala Ile Ala Thr His Val His Thr Glu  
 275 280 285  
 35 Met His Arg Ile Leu Ala Ser Ala Gly Ala Asn Gly Pro Gly Pro Glu  
 290 295 300  
 Leu Leu Phe Tyr His Cys Glu Pro Pro Gly Gly Ala Val Leu Tyr Pro  
 305 310 315 320  
 Phe Phe Leu Leu Asn Lys Gln Lys Thr Pro Ala Phe Glu Tyr Phe Ile  
 325 330 335  
 40 Lys Lys Phe Asn Ser Gly Gly Val Met Ala Ser Gln Glu Leu Val Ser  
 340 345 350  
 Val Thr Val Arg Leu Gln Thr Asp Pro Val Glu Tyr Leu Ser Glu Gln  
 245

355                      360                      365  
 Leu Asn Asn Leu Ile Glu Thr Val Ser Pro Asn Thr Asp Val Arg Met  
 370                      375                      380  
 Tyr Ser Gly Lys Arg Asn Gly Ala Ala Asp Asp Leu Met Val Ala Val  
 5    385                      390                      395                      400  
 Ile Met Ala Ile Tyr Leu Ala Ala Pro Thr Gly Ile Pro Pro Ala Phe  
 405                      410                      415  
 Phe Pro Ile Thr Arg Thr Ser  
 420

10

## (2) INFORMATION FOR SEQ ID NO:43:

## (i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 355 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Val Leu Leu Ser Pro Ala Pro Pro Pro Leu Pro His Gly Arg Cys Pro  
 1                      5                      10                      15  
 25    Pro Ser Leu Phe His His Arg Pro Gly Cys Val Ser Gly Pro Pro Ala  
 20                      25                      30  
 Pro Pro Arg Ser Gly Val Ser Arg Pro Gly Ala Met Ile Thr Asp Cys  
 35                      40                      45  
 Phe Glu Ala Asp Ile Ala Ile Pro Ser Gly Ile Ser Arg Pro Asp Ala  
 30    50                      55                      60  
 Ala Ala Leu Gln Arg Cys Glu Gly Arg Val Val Phe Leu Pro Thr Ile  
 65                      70                      75                      80  
 Arg Arg Gln Leu Ala Asp Val Ala His Glu Ser Phe Val Ser Gly Gly  
 85                      90                      95  
 35    Val Ser Pro Asp Thr Leu Gly Leu Leu Leu Ala Tyr Arg Arg Arg Phe  
 100                      105                      110  
 Pro Ala Val Ile Thr Arg Val Leu Pro Thr Arg Ile Val Ala Cys Pro  
 115                      120                      125  
 Val Asp Leu Gly Leu Thr His Ala Gly Thr Val Asn Leu Arg Asn Thr  
 40    130                      135                      140  
 Ser Pro Val Asp Leu Cys Asn Gly Asp Pro Val Ser Leu Val Pro Pro  
 145                      150                      155                      160  
 Val Phe Glu Gly Gln Ala Thr Asp Val Arg Leu Glu Ser Leu Asp Leu

165 170 175  
 Thr Leu Arg Phe Pro Val Pro Leu Pro Thr Pro Leu Ala Arg Glu Ile  
 180 185 190  
 Val Ala Arg Leu Val Arg Ile Arg Asp Leu Asn Pro Asp Pro Arg Thr  
 5 195 200 205  
 Pro Gly Glu Leu Pro Asp Leu Asn Val Leu Tyr Tyr Asn Gly Ala Arg  
 210 215 220  
 Leu Ser Leu Val Ala Asp Val Gln Gln Leu Ala Ser Val Asn Thr Glu  
 225 230 235 240  
 10 Leu Arg Ser Leu Val Leu Asn Met Val Tyr Ser Ile Thr Glu Gly Thr  
 245 250 255  
 Thr Leu Ile Leu Thr Leu Ile Pro Arg Leu Leu Ala Leu Ser Ala Gln  
 260 265 270  
 Asp Gly Tyr Val Asn Ala Leu Leu Gln Met Gln Ser Val Thr Arg Glu  
 15 275 280 285  
 Ala Ala Gln Leu Ile His Pro Glu Ala Pro Met Leu Met Gln Asp Gly  
 290 295 300  
 Glu Arg Arg Leu Pro Leu Tyr Glu Ala Leu Val Ala Trp Leu Ala His  
 305 310 315 320  
 20 Ala Gly Gln Leu Gly Asp Ile Leu Ala Pro Ala Val Arg Val Cys Thr  
 325 330 335  
 Phe Asp Gly Ala Ala Val Val Gln Ser Gly Asp Met Ala Pro Val Ile  
 340 345 350  
 Arg Tyr Pro  
 25 355

## (2) INFORMATION FOR SEQ ID NO:44:

## (i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 1382 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

40 Val Trp Glu Gly Leu Gly Leu Pro Glu Leu Gly Leu Met Glu Pro Ala  
 1 5 10 15  
 Asn Pro Pro Arg Asn Pro Met Ala Ala Pro Ala Arg Asp Pro Pro Gly  
 20 25 30  
 Tyr Arg Tyr Ala Ala Ala Met Val Pro Thr Gly Ser Ile Leu Ser Thr  
 247

	35	40	45
	Ile Glu Val Ala Ser His Arg Arg Leu Phe Asp Phe Phe Ala Arg Val		
	50	55	60
5	Arg Ser Asp Glu Asn Ser Leu Tyr Asp Val Glu Phe Asp Ala Leu Leu		
	65	70	75
	Gly Ser Tyr Cys Asn Thr Leu Ser Leu Val Arg Phe Leu Glu Leu Gly		
	85	90	95
	Leu Ser Val Ala Cys Val Cys Thr Lys Phe Pro Glu Leu Ala Tyr Met		
	100	105	110
10	Asn Glu Gly Arg Val Gln Phe Glu Val His Gln Pro Leu Ile Ala Arg		
	115	120	125
	Asp Gly Pro His Pro Val Glu Gln Pro Val His Asn Tyr Met Thr Lys		
	130	135	140
	Val Ile Asp Arg Arg Ala Leu Asn Ala Ala Phe Ser Leu Ala Thr Glu		
15	145	150	155
	Ala Ile Ala Leu Leu Thr Gly Glu Ala Leu Asp Gly Thr Gly Ile Ser		
	165	170	175
	Leu His Arg Gln Leu Arg Ala Ile Gln Gln Leu Ala Arg Asn Val Gln		
	180	185	190
20	Ala Val Leu Gly Ala Phe Glu Arg Gly Thr Ala Asp Gln Met Leu His		
	195	200	205
	Val Leu Leu Glu Lys Ala Pro Pro Leu Ala Leu Leu Leu Pro Met Gln		
	210	215	220
	Arg Tyr Leu Asp Asn Gly Arg Leu Ala Thr Arg Val Ala Arg Ala Thr		
25	225	230	235
	Leu Val Ala Glu Leu Lys Arg Ser Phe Cys Asp Thr Ser Phe Phe Leu		
	245	250	255
	Gly Lys Ala Gly His Arg Arg Glu Ala Ile Glu Ala Trp Leu Val Asp		
	260	265	270
30	Leu Thr Thr Ala Thr Gln Pro Ser Val Ala Val Pro Arg Leu Thr His		
	275	280	285
	Ala Asp Thr Arg Gly Arg Pro Val Asp Gly Val Leu Val Thr Thr Ala		
	290	295	300
	Ala Ile Lys Gln Arg Leu Leu Gln Ser Phe Leu Lys Val Glu Asp Thr		
35	305	310	315
	Glu Ala Asp Val Pro Val Thr Tyr Gly Glu Met Val Leu Asn Gly Ala		
	325	330	335
	Asn Leu Val Thr Ala Leu Val Met Gly Lys Ala Val Arg Ser Leu Asp		
	340	345	350
40	Asp Val Gly Arg His Leu Leu Glu Met Gln Glu Glu Gln Leu Glu Ala		
	355	360	365
	Asn Arg Glu Thr Leu Asp Glu Leu Glu Ser Ala Pro Gln Thr Thr Arg		
	370	375	380

Val Arg Ala Asp Leu Val Ala Ile Gly Asp Arg Leu Val Phe Leu Glu  
 385 390 395 400  
 Ala Leu Glu Lys Arg Ile Tyr Ala Ala Thr Asn Val Pro Tyr Pro Leu  
 405 410 415  
 5 Val Gly Ala Met Asp Leu Thr Phe Val Leu Pro Leu Gly Leu Phe Asn  
 420 425 430  
 Pro Ala Met Glu Arg Phe Ala Ala His Ala Gly Asp Leu Val Pro Ala  
 435 440 445  
 Pro Gly His Pro Glu Pro Arg Ala Phe Pro Pro Arg Gln Leu Phe Phe  
 10 450 455 460  
 Trp Gly Lys Asp His Gln Val Leu Arg Leu Ser Met Glu Asn Ala Val  
 465 470 475 480  
 Gly Thr Val Cys His Pro Ser Leu Met Asn Ile Asp Ala Ala Val Gly  
 485 490 495  
 15 Gly Val Asn His Asp Pro Val Glu Ala Ala Asn Pro Tyr Gly Ala Tyr  
 500 505 510  
 Val Ala Ala Pro Ala Gly Pro Gly Ala Asp Met Gln Gln Arg Phe Leu  
 515 520 525  
 Asn Ala Trp Arg Gln Arg Leu Ala His Gly Arg Val Arg Trp Val Ala  
 20 530 535 540  
 Glu Cys Gln Met Thr Ala Glu Gln Phe Met Gln Pro Asp Asn Ala Asn  
 545 550 555 560  
 Leu Ala Leu Glu Leu His Pro Ala Phe Asp Phe Phe Ala Gly Val Ala  
 565 570 575  
 25 Asp Val Glu Leu Pro Gly Gly Glu Val Pro Pro Ala Gly Pro Gly Ala  
 580 585 590  
 Ile Gln Ala Thr Trp Arg Val Val Asn Gly Asn Leu Pro Leu Ala Leu  
 595 600 605  
 Cys Pro Val Ala Phe Arg Asp Arg Leu Glu Leu Gly Val Gly Arg His  
 30 610 615 620  
 Ala Met Ala Pro Ala Thr Ile Ala Ala Val Arg Gly Ala Phe Glu Asp  
 625 630 635 640  
 Arg Ser Tyr Pro Ala Val Phe Tyr Leu Leu Gln Ala Ala Ile His Gly  
 645 650 655  
 35 Ser Glu His Val Phe Cys Ala Arg Leu Val Thr Gln Cys Ile Thr Ser  
 660 665 670  
 Tyr Trp Asn Asn Thr Arg Cys Ala Ala Phe Val Asn Asp Tyr Ser Leu  
 675 680 685  
 Val Ser Tyr Ile Val Thr Tyr Leu Gly Gly Asp Leu Pro Glu Glu Cys  
 40 690 695 700  
 Met Ala Val Tyr Arg Asp Leu Val Ala His Val Glu Ala Gln Leu Val  
 705 710 715 720  
 Asp Asp Phe Thr Leu Pro Gly Pro Glu Leu Gly Gly Gln Ala Gln Ala



		725		730		735	
		Glu Leu Asn His Leu Met Arg Asp Pro Ala Leu Leu Pro Pro Leu Val					
		740		745		750	
		Trp Asp Cys Asp Gly Leu Met Arg His Ala Ala Leu Asp Arg His Arg					
5		755		760		765	
		Asp Cys Arg Ile Asp Ala Gly Gly His Glu Pro Val Tyr Ala Ala Ala					
		770		775		780	
		Cys Asn Val Ala Thr Ala Asp Phe Asn Arg Asn Asp Gly Arg Leu Leu					
		785		790		795	800
10		His Asn Thr Gln Ala Arg Ala Ala Asp Ala Ala Asp Asp Arg Pro His					
		805		810		815	
		Arg Pro Ala Asp Trp Thr Val His His Lys Ile Tyr Tyr Tyr Val Leu					
		820		825		830	
		Val Pro Ala Phe Ser Arg Gly Arg Cys Cys Thr Ala Gly Val Arg Phe					
15		835		840		845	
		Asp Arg Val Tyr Ala Thr Leu Gln Asn Met Val Val Pro Glu Ile Ala					
		850		855		860	
		Pro Gly Glu Glu Cys Pro Ser Asp Pro Val Thr Asp Pro Ala His Pro					
		865		870		875	880
20		Leu His Pro Ala Asn Leu Val Ala Asn Thr Val Asn Ala Met Phe His					
		885		890		895	
		Asn Gly Arg Val Val Val Asp Gly Pro Ala Met Leu Thr Leu Gln Val					
		900		905		910	
		Leu Ala His Asn Met Ala Glu Arg Thr Thr Ala Leu Leu Cys Ser Ala					
25		915		920		925	
		Ala Pro Asp Ala Gly Ala Asn Thr Ala Ser Thr Ala Asn Met Arg Ile					
		930		935		940	
		Phe Asp Gly Ala Leu His Ala Gly Val Leu Leu Met Ala Pro Gln His					
		945		950		955	960
30		Leu Asp His Thr Ile Gln Asn Gly Glu Tyr Phe Tyr Val Leu Pro Val					
		965		970		975	
		His Ala Leu Phe Ala Gly Ala Asp His Val Ala Asn Ala Pro Asn Phe					
		980		985		990	
		Pro Pro Ala Leu Arg Asp Leu Ala Arg His Val Pro Leu Val Pro Pro					
35		995		1000		1005	
		Ala Leu Gly Ala Asn Tyr Phe Ser Ser Ile Arg Gln Pro Val Val Gln					
		1010		1015		1020	
		His Ala Arg Glu Ser Ala Ala Gly Glu Asn Ala Leu Thr Tyr Ala Leu					
		1025		1030		1035	104
40		Met Ala Gly Tyr Phe Lys Met Ser Pro Val Tyr His Gln Leu Lys Thr					
		1045		1050		1055	
		Gly Leu His Pro Gly Phe Gly Phe Thr Val Val Arg Gln Asp Arg Phe					
		1060		1065		1070	

Val Thr Glu Asn Val Leu Phe Ser Ala Ser Glu Ala Tyr Phe Leu Gly  
 1075 1080 1085  
 Gln Leu Gln Val Ala Arg His Glu Thr Gly Gly Gly Val Ser Phe Thr  
 1090 1095 1100  
 5 Leu Thr Gln Pro Arg Gly Asn Val Asp Leu Gly Val Gly Tyr Thr Ala  
 1105 1110 1115 112  
 Val Ala Ala Thr Ala Thr Val Arg Asn Pro Val Thr Asp Met Gly Asn  
 1125 1130 1135  
 Leu Pro Gln Asn Phe Tyr Leu Gly Arg Gly Ala Pro Pro Leu Leu Asn  
 10 1140 1145 1150  
 Asn Ala Ala Ala Val Tyr Leu Arg Asn Ala Val Val Ala Gly Asn Arg  
 1155 1160 1165  
 Leu Gly Pro Ala Gln Pro Leu Pro Val Phe Gly Cys Ala Gln Val Pro  
 1170 1175 1180  
 15 Arg Arg Ala Gly Met Asp His Gly Gln Asp Ala Val Cys Glu Phe Ile  
 1185 1190 1195 120  
 Ala Thr Pro Val Ala Thr Asp Ile Asn Tyr Phe Arg Arg Pro Cys Asn  
 1205 1210 1215  
 Pro Arg Gly Arg Ala Ala Gly Gly Val Tyr Ala Gly Asp Lys Glu Gly  
 20 1220 1225 1230  
 Asp Val Ile Ala Leu Met Tyr Asp His Gly Gln Ser Asp Pro Ala Arg  
 1235 1240 1245  
 Pro Phe Ala Ala Thr Ala Asn Pro Trp Ala Ser Gln Arg Phe Ser Tyr  
 1250 1255 1260  
 25 Gly Asp Leu Leu Tyr Asn Gly Ala Tyr His Leu Asn Gly Asp Val Leu  
 1265 1270 1275 128  
 Ser Pro Cys Phe Lys Phe Phe Thr Ala Ala Asp Ile Thr Ala Lys His  
 1285 1290 1295  
 Arg Cys Leu Glu Arg Leu Ile Val Glu Thr Gly Ser Ala Val Ser Thr  
 30 1300 1305 1310  
 Ala Thr Ala Ala Ser Asp Val Gln Phe Lys Arg Pro Pro Gly Cys Arg  
 1315 1320 1325  
 Glu Leu Val Glu Asp Pro Cys Gly Leu Phe Gln Glu Ala Tyr Pro Ile  
 1330 1335 1340  
 35 Thr Cys Ala Ser Asp Pro Ala Leu Leu Arg Ser Ala Arg Asp Gly Glu  
 1345 1350 1355 136  
 Ala His Ala Arg Glu Thr His Phe Thr Gln Tyr Leu Ile Tyr Asp Asp  
 1365 1370 1375  
 Leu Lys Gly Leu Ser Leu  
 40 1380

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 222 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

10

	Met	Thr	Met	Arg	Asp	Asp	Val	Pro	Leu	Leu	Asp	Arg	Glu	Leu	Val	Tyr
	1				5				10						15	
	Glu	Ala	Ala	Cys	Gly	Gly	Glu	Asp	Gly	Glu	Leu	Pro	Leu	Asp	Glu	Gln
				20					25					30		
15	Phe	Ser	Leu	Ser	Ser	Tyr	Gly	Thr	Ser	Asp	Phe	Phe	Val	Ser	Ser	Ala
			35					40					45			
	Tyr	Ser	Arg	Leu	Pro	Pro	His	Thr	Gln	Pro	Val	Phe	Ser	Lys	Arg	Val
	50						55					60				
20	Val	Met	Phe	Ala	Trp	Ser	Phe	Leu	Val	Leu	Lys	Pro	Leu	Glu	Leu	Val
	65					70					75				80	
	Ala	Ala	Gly	Met	Tyr	Tyr	Gly	Trp	Thr	Gly	Arg	Ala	Val	Ala	Pro	Ala
				85						90					95	
	Cys	Ile	Ile	Ala	Ala	Val	Leu	Ala	Tyr	Tyr	Val	Thr	Trp	Leu	Ala	Arg
				100					105					110		
25	Ala	Leu	Leu	Leu	Tyr	Val	Asn	Ile	Lys	Arg	Asp	Arg	Leu	Pro	Leu	Ser
			115					120					125			
	Pro	Pro	Val	Phe	Trp	Gly	Leu	Cys	Val	Ile	Met	Gly	Gly	Ala	Ala	Leu
	130						135					140				
30	Cys	Ala	Leu	Val	Ala	Ala	Ala	His	Glu	Thr	Phe	Ser	Pro	Asp	Gly	Leu
	145					150					155				160	
	Phe	His	Trp	Ile	Thr	Ala	Ser	Gln	Leu	Leu	Pro	Arg	Thr	Asp	Pro	Leu
				165						170					175	
	Arg	Ala	Arg	Ser	Leu	Gly	Ile	Ala	Cys	Ala	Ala	Gly	Ala	Ala	Met	Trp
				180					185				190			
35	Val	Ala	Ala	Ala	Asp	Cys	Phe	Ala	Ala	Phe	Thr	Asn	Phe	Phe	Leu	Ala
			195					200				205				
	Arg	Phe	Trp	Thr	Arg	Ala	Ile	Leu	Lys	Ala	Pro	Val	Ala	Phe		
	210						215					220				

40

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 627 amino acids

(B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

```

Val Gly Arg Gln Gly Glu Arg Trp Val Gly Gly Gly Asn Glu Lys Asn
10   1           5           10           15
Thr Gln Arg Ala Thr Ser Gly Met Arg Pro Glu Leu Ser Leu Lys Gly
      20           25           30
Arg Pro Cys Val Thr Glu Ala Val Val Cys Pro Ser Thr Asp Ala Ala
      35           40           45
15 Ile His Ser Gly Gly Ser Ser Ser Val Arg Pro Gln Pro Tyr Ala Arg
      50           55           60
Ala Ala Arg Ala Arg Ala Thr His Gly Ser Arg Ser Arg His Arg Gln
      65           70           75           80
Pro Leu Leu Pro Pro Pro Ser Ser His His Pro Thr Ile Pro Pro Pro
20           85           90           95
Pro Ser Pro Pro Arg Gly Ser Pro Ala Met Glu Leu Thr Tyr Ala Thr
      100          105          110
Thr Leu His His Arg Asp Val Val Phe Tyr Val Thr Ala Asp Arg Asn
      115          120          125
25 Arg Ala Tyr Phe Val Cys Gly Gly Ser Val Tyr Ser Val Gly Arg Pro
      130          135          140
Arg Asp Ser Gln Pro Gly Glu Ile Ala Lys Phe Gly Leu Val Val Arg
      145          150          155          160
Gly Thr Gly Pro Lys Asp Arg Met Val Ala Asn Tyr Val Arg Ser Glu
30           165          170          175
Leu Arg Gln Arg Gly Leu Arg Glu Val Arg Pro Val Gly Glu Asp Glu
      180          185          190
Val Phe Leu Asp Ser Val Cys Leu Leu Asn Pro Asn Val Ser Ser Asp
      195          200          205
35 Val Ile Asn Thr Asn Asp Val Glu Val Leu Asp Glu Cys Leu Ala Glu
      210          215          220
Tyr Cys Thr Ser Leu Gln Thr Ser Pro Gly Val Leu Val Thr Gly Val
      225          230          235          240
Arg Val Arg Ala Arg Asp Arg Val Ile Glu Leu Phe Glu His Pro Ala
40           245          250          255
Ile Val Asn Ile Ser Ser Arg Phe Ala Tyr Thr Pro Ser Pro Tyr Val
      260          265          270
Phe Ala Gln Ala His Leu Pro Arg Leu Pro Ser Ser Leu Glu Pro Leu
      253

```

	275		280		285
	Val Ser Gly Leu Phe Asp Gly Ile Pro Ala Pro Arg Gln Pro Leu Asp				
	290		295		300
	Ala Arg Asp Arg Arg Thr Asp Val Val Ile Thr Gly Thr Arg Ala Pro				
5	305		310		315
	Arg Pro Met Ala Gly Thr Gly Ala Gly Gly Ala Gly Ala Lys Arg Ala				
		325		330	
	Thr Val Ser Glu Phe Val Gln Val Lys His Ile Asp Arg Val Val Ser				
		340		345	
10	Pro Ser Val Ser Ser Ala Pro Pro Pro Ser Ala Pro Asp Ala Ser Leu				
		355		360	
	Pro Pro Pro Gly Leu Gln Glu Ala Ala Pro Pro Gly Pro Pro Leu Arg				
		370		375	
	Glu Leu Trp Trp Val Phe Tyr Ala Gly Asp Arg Ala Leu Glu Glu Pro				
15	385		390		395
	His Ala Glu Ser Gly Leu Thr Arg Glu Glu Val Arg Ala Val His Gly				
		405		410	
	Phe Arg Glu Gln Ala Trp Lys Leu Phe Gly Ser Val Gly Ala Pro Arg				
		420		425	
20	Ala Phe Leu Gly Ala Ala Leu Ser Pro Thr Gln Lys Leu Ala Val Tyr				
		435		440	
	Tyr Tyr Leu Ile His Arg Glu Arg Arg Met Ser Pro Phe Pro Ala Leu				
		450		455	
	Val Arg Leu Val Gly Arg Tyr Ile Gln Arg His Gly Val Pro Ala Pro				
25	465		470		475
	Asp Glu Pro Thr Leu Ala Asp Ala Met Asn Gly Leu Phe Arg Asp Ala				
		485		490	
	Ala Gly Thr Val Ala Glu Gln Leu Leu Met Phe Asp Leu Leu Pro Pro				
		500		505	
30	Lys Asp Val Pro Val Gly Ser Asp Ala Arg Ala Asp Ser Ala Ala Leu				
		515		520	
	Leu Arg Phe Val Asp Ser Gln Arg Leu Thr Pro Gly Gly Ser Val Ser				
		530		535	
	Pro Glu His Val Met Tyr Leu Gly Ala Phe Leu Gly Val Leu Tyr Ala				
35	545		550		555
	Gly His Gly Arg Leu Ala Ala Ala Thr His Thr Ala Arg Leu Thr Gly				
		565		570	
	Val Thr Ser Leu Val Leu Thr Val Gly Asp Val Asp Arg Met Ser Ala				
		580		585	
40	Phe Asp Arg Gly Pro Ala Gly Ala Ala Gly Arg Thr Arg Thr Ala Gly				
		595		600	
	Tyr Leu Asp Ala Leu Leu Thr Val Cys Leu Ala Arg Ala Gln His Gly				
		610		615	
				620	

Gln Ser Val  
625

## (2) INFORMATION FOR SEQ ID NO:47:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 592 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
10 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

15

## (x) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Val Tyr Leu Ser Pro Ser Ala Leu Lys Trp Pro Val Gly Val Trp Thr  
1 5 10 15  
Thr Gly Gly Leu Ala Phe Gly Cys Asp Ala Ala Leu Val Arg Ala Arg  
20 25 30  
Tyr Gly Lys Gly Phe Met Gly Val Val Ile Ser Met Arg Asp Ser Pro  
35 40 45  
Pro Ala Glu Ile Ile Val Val Pro Ala Asp Lys Thr Leu Ala Arg Val  
50 55 60  
Gly Asn Pro Thr Asp Glu Asn Ala Pro Ala Val Leu Pro Gly Pro Pro  
25 65 70 75 80  
Ala Gly Pro Arg Tyr Arg Val Phe Val Leu Gly Ala Pro Thr Pro Ala  
85 90 95  
Asp Asn Gly Ser Ala Leu Asp Ala Leu Arg Arg Val Ala Gly Tyr Pro  
100 105 110  
30 Glu Glu Ser Thr Asn Tyr Ala Gln Tyr Met Ser Arg Ala Tyr Ala Glu  
115 120 125  
Phe Leu Gly Glu Asp Pro Gly Ser Gly Thr Asp Ala Arg Pro Ser Leu  
130 135 140  
Phe Trp Arg Leu Ala Gly Leu Leu Ala Ser Ser Gly Phe Ala Phe Val  
35 145 150 155 160  
Asn Ala Ala His Ala His Asp Ala Ile Arg Leu Ser Asp Leu Leu Gly  
165 170 175  
Phe Leu Ala His Ser Arg Val Leu Ala Gly Leu Ala Arg Ala Ala Gly  
180 185 190  
40 Cys Ala Ala Asp Ser Val Phe Leu Asn Val Ser Val Leu Asp Pro Ala  
195 200 205  
Ala Arg Leu Arg Leu Glu Ala Arg Leu Gly His Leu Val Ala Ala Ile  
210 215 220

255

Arg Glu Gln Ser Leu Ala Ala His Ala Leu Gly Tyr Gln Leu Ala Phe  
 225 230 235 240  
 Val Leu Asp Ser Pro Ala Ala Tyr Gly Ala Val Ala Pro Ser Ala Ala  
 245 250 255  
 5 Arg Leu Ile Asp Ala Leu Tyr Ala Glu Phe Leu Gly Gly Arg Ala Leu  
 260 265 270  
 Thr Ala Pro Met Val Arg Arg Ala Leu Phe Tyr Ala Thr Ala Val Leu  
 275 280 285  
 Arg Ala Pro Phe Leu Ala Gly Ala Pro Ser Ala Glu Gln Arg Glu Arg  
 10 290 295 300  
 Ala Arg Arg Gly Leu Leu Ile Thr Thr Ala Leu Cys Thr Ser Asp Val  
 305 310 315 320  
 Ala Ala Ala Thr His Ala Asp Leu Arg Ala Ala Arg Thr Asp His Gln  
 325 330 335  
 15 Lys Asn Leu Phe Trp Leu Pro Asp His Phe Ser Pro Cys Ala Ala Ser  
 340 345 350  
 Leu Arg Phe Asp Leu Ala Glu Gly Gly Phe Ile Leu Asp Ala Met Ala  
 355 360 365  
 Thr Arg Ser Asp Ile Pro Ala Asp Val Met Ala Gln Gln Thr Arg Gly  
 20 370 375 380  
 Val Ala Ser Val Leu Thr Arg Trp Ala His Tyr Asn Ala Leu Ile Arg  
 385 390 395 400  
 Ala Phe Val Pro Glu Ala Thr His Gln Cys Ser Gly Pro Ser His Asn  
 405 410 415  
 25 Ala Glu Pro Arg Ile Leu Val Pro Ile Thr His Asn Ala Ser Tyr Val  
 420 425 430  
 Val Thr His Thr Pro Leu Pro Arg Gly Ile Gly Tyr Lys Leu Thr Gly  
 435 440 445  
 Val Asp Val Arg Arg Pro Leu Phe Ile Thr Tyr Leu Thr Ala Thr Cys  
 30 450 455 460  
 Glu Gly His Ala Arg Glu Ile Glu Pro Lys Arg Leu Val Arg Thr Glu  
 465 470 475 480  
 Asn Arg Arg Asp Leu Gly Leu Val Gly Ala Val Phe Leu Arg Tyr Thr  
 485 490 495  
 35 Pro Ala Gly Glu Val Met Ser Val Leu Leu Val Asp Thr Asp Ala Thr  
 500 505 510  
 Gln Gln Gln Leu Ala Gln Gly Pro Val Ala Gly Thr Pro Asn Val Phe  
 515 520 525  
 Ser Ser Asp Val Pro Ser Val Leu Leu Phe Pro Asn Gly Thr Val Ile  
 40 530 535 540  
 His Leu Leu Ala Phe Asp Thr Leu Pro Ile Ala Thr Ile Ala Pro Gly  
 545 550 555 560  
 Phe Leu Ala Ala Ser Ala Leu Gly Val Val Met Ile Thr Ala Ala Gly  
 256

565 570 575  
 Ile Leu Arg Val Val Arg Thr Cys Val Pro Phe Leu Trp Arg Arg Glu  
 580 585 590

5 (2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 315 amino acids

(B) TYPE: amino acid

10 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Val Ser Ile Ser Ala Gly Val Arg Gly Gln Gly Trp His Arg Ile Ser  
 1 5 10 15  
 Thr Pro Pro Lys Asn Gly Ala Gly Arg Ser Val Leu Val Phe Gly Leu  
 20 20 25 30  
 Val Leu Pro Leu Cys Phe Tyr Pro His Pro Thr Pro Ser Phe Gly Pro  
 35 40 45  
 Arg Leu Arg Gln Gln Arg Ala Ser Asp Ser Leu Arg Gly Ala Glu Pro  
 50 55 60  
 25 Leu Trp Ala Val Gly Thr Asp Thr Pro Pro Ser Ala Asp Trp Gln Pro  
 65 70 75 80  
 Gly Arg Thr Thr Met Gly Pro Gly Leu Trp Val Val Met Gly Val Leu  
 85 90 95  
 Val Gly Val Ala Gly Gly His Asp Thr Tyr Trp Thr Glu Gln Ile Asp  
 30 100 105 110  
 Pro Trp Phe Leu His Gly Leu Gly Leu Ala Arg Thr Tyr Trp Arg Asp  
 115 120 125  
 Thr Asn Thr Gly Arg Leu Trp Leu Pro Asn Thr Pro Asp Ala Ser Asp  
 130 135 140  
 35 Pro Gln Arg Gly Arg Leu Ala Pro Pro Gly Glu Leu Asn Leu Thr Thr  
 145 150 155 160  
 Ala Ser Val Pro Met Leu Arg Trp Tyr Ala Glu Arg Phe Cys Phe Val  
 165 170 175  
 Leu Val Thr Thr Ala Glu Phe Pro Arg Asp Pro Gly Gln Leu Leu Tyr  
 40 180 185 190  
 Ile Pro Lys Thr Tyr Leu Leu Gly Arg Pro Arg Asn Ala Ser Leu Pro  
 195 200 205  
 Glu Leu Pro Glu Ala Gly Pro Thr Ser Arg Pro Pro Ala Glu Val Thr  
 257



210                      215                      220  
 Gln Leu Lys Gly Leu Ser His Asn Pro Gly Ala Ser Ala Leu Leu Arg  
 225                      230                      235                      240  
 Ser Arg Ala Trp Val Thr Phe Ala Ala Ala Pro Asp Arg Glu Gly Leu  
 5                      245                      250                      255  
 Thr Phe Pro Arg Gly Asp Asp Gly Ala Thr Glu Arg His Pro Asp Gly  
 260                      265                      270  
 Arg Arg Asn Ala Pro Pro Pro Gly Pro Pro Ala Gly Thr Pro Arg His  
 275                      280                      285  
 10 Pro Thr Thr Asn Leu Ser Ile Ala His Leu His Asn Ala Ser Val Thr  
 290                      295                      300  
 Trp Leu Ala Arg Leu Leu Arg Thr Pro Gly Arg  
 305                      310                      315

15                      (2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 370 amino acids

(B) TYPE: amino acid

20                      (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

25                      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Met Ala Ser His Ala Gly Gln Gln His Ala Pro Ala Phe Gly Gln Ala  
 1                      5                      10                      15  
 Ala Arg Ala Ser Gly Pro Thr Asp Gly Arg Ala Ala Ser Arg Pro Ser  
 30                      20                      25                      30  
 His Arg Gln Gly Ala Ser Asp Pro Glu Leu Pro Thr Leu Leu Arg Val  
 35                      40                      45  
 Tyr Ile Asp Gly Pro His Gly Val Gly Lys Thr Thr Thr Ser Ala Gln  
 50                      55                      60  
 35 Leu Met Glu Ala Leu Gly Pro Arg Asp Asn Ile Val Tyr Val Pro Glu  
 65                      70                      75                      80  
 Pro Met Thr Tyr Trp Gln Val Leu Gly Ala Ser Glu Thr Leu Thr Asn  
 85                      90                      95  
 Ile Tyr Asn Thr Gln His Arg Leu Asp Arg Gly Glu Ile Ser Ala Gly  
 40                      100                      105                      110  
 Glu Ala Ala Val Val Met Thr Ser Ala Gln Ile Thr Met Ser Thr Pro  
 115                      120                      125  
 Tyr Ala Ala Thr Asp Ala Val Leu Ala Pro His Ile Gly Gly Glu Ala  
 258

	130	135	140
	Val Gly Pro Gln Ala Pro Pro Pro Ala Leu Thr Leu Val Phe Asp Arg		
	145	150	155
	His Pro Ile Ala Ser Leu Leu Cys Tyr Pro Ala Ala Arg Tyr Leu Met		160
5	165	170	175
	Gly Ser Met Thr Pro Gln Ala Val Leu Ala Phe Val Met Pro Pro Thr		
	180	185	190
	Ala Pro Gly Thr Asn Leu Val Leu Gly Val Leu Pro Glu Ala Glu His		
	195	200	205
10	Ala Asp Arg Leu Ala Arg Arg Gln Arg Pro Gly Glu Arg Leu Asp Leu		
	210	215	220
	Ala Met Leu Ser Ala Ile Arg Arg Val Tyr Asp Leu Leu Ala Asn Thr		
	225	230	235
	Val Arg Tyr Leu Gln Arg Gly Gly Arg Trp Arg Glu Asp Trp Gly Arg		240
15	245	250	255
	Leu Thr Gly Val Ala Ala Ala Thr Pro Arg Pro Asp Pro Glu Asp Gly		
	260	265	270
	Ala Gly Ser Leu Pro Arg Ile Glu Asp Thr Leu Phe Ala Leu Phe Arg		
	275	280	285
20	Val Pro Glu Leu Leu Ala Pro Asn Gly Asp Leu Tyr His Ile Phe Ala		
	290	295	300
	Trp Val Leu Asp Val Leu Ala Asp Arg Leu Leu Pro Met His Leu Phe		
	305	310	315
	Val Leu Asp Tyr Asp Gln Ser Pro Val Gly Cys Arg Asp Ala Leu Leu		320
25	325	330	335
	Arg Leu Thr Ala Gly Met Ile Pro Thr Arg Val Thr Thr Ala Gly Ser		
	340	345	350
	Ile Ala Glu Ile Arg Asp Leu Ala Arg Thr Phe Ala Arg Glu Val Gly		
	355	360	365
30	Gly Val		
	370		

## (2) INFORMATION FOR SEQ ID NO:50:

## 35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 352 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Val Leu Arg Val Val Asp Val Arg Gln Gly Leu Gly Gly Pro Gln His  
 1 5 10 15  
 Leu Pro Val Ser His Arg Leu Gly Asp Val Asp Asp Ile Val Ala Arg  
 5 20 25 30  
 Pro Gln Gly Leu His Gln Leu Arg Gly Gly Gly Gly Leu Pro His Pro  
 35 40 45  
 Val Gly Ser Val Tyr Ile Asn Pro Gln Gln Arg Gly Gln Leu Arg Ile  
 50 55 60  
 10 Pro Ala Gly Phe Gly Gly Pro Leu Ala Met Ala Arg Thr Gly Arg Arg  
 65 70 75 80  
 Ala Ala Val Gly Arg Pro Ala Arg Thr Ser Ser Leu Thr Glu Arg Arg  
 85 90 95  
 Arg Val Leu Leu Ala Gly Val Arg Ser His Thr Arg Phe Tyr Lys Ala  
 15 100 105 110  
 Phe Ala Arg Glu Val Arg Glu Phe Asn Ala Thr Arg Ile Cys Gly Thr  
 115 120 125  
 Leu Leu Thr Leu Met Ser Gly Ser Leu Gln Gly Arg Ser Leu Phe Glu  
 130 135 140  
 20 Ala Thr Arg Val Thr Leu Ile Cys Glu Val Asp Leu Gly Pro Arg Arg  
 145 150 155 160  
 Pro Asp Cys Ile Cys Val Phe Glu Phe Ala Asn Asp Lys Thr Leu Gly  
 165 170 175  
 Gly Val Cys Val Ile Leu Lys Thr Cys Lys Ser Ile Ser Ser Gly Asp  
 25 180 185 190  
 Thr Ala Ser Lys Arg Glu Gln Arg Thr Thr Gly Met Lys Gln Leu Arg  
 195 200 205  
 His Ser Leu Lys Leu Leu Gln Ser Leu Ala Pro Pro Gly Asp Lys Val  
 210 215 220  
 30 Val Tyr Leu Cys Pro Ile Leu Val Phe Val Ala Gln Arg Thr Leu Arg  
 225 230 235 240  
 Val Ser Arg Val Thr Arg Leu Val Pro Gln Lys Ile Ser Gly Asn Ile  
 245 250 255  
 Thr Ala Ala Val Arg Met Leu Gln Ser Leu Ser Thr Tyr Ala Val Pro  
 35 260 265 270  
 Pro Glu Pro Gln Thr Arg Arg Ser Arg Arg Arg Val Ala Ala Thr Ala  
 275 280 285  
 Arg Pro Gln Arg Pro Pro Ser Pro Thr Arg Asp Pro Glu Gly Thr Ala  
 290 295 300  
 40 Gly His Pro Ala Pro Pro Glu Ser Asp Pro Pro Ser Pro Gly Val Val  
 305 310 315 320  
 Gly Val Ala Ala Glu Gly Gly Gly Val Leu Gln Lys Ile Ala Ala Leu  
 325 330 335

Phe Cys Val Pro Val Ala Ala Lys Ser Arg Pro Arg Thr Lys Thr Glu  
 340 345 350

(2) INFORMATION FOR SEQ ID NO:51:

5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 514 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

15

Met Asp Pro Tyr Tyr Pro Phe Asp Ala Leu Asp Val Trp Glu His Arg  
 1 5 10 15  
 Arg Phe Ile Val Ala Asp Ser Arg Ser Phe Ile Thr Pro Glu Phe Pro  
 20 25 30  
 Arg Asp Phe Trp Met Leu Pro Val Phe Asn Ile Pro Arg Glu Thr Ala  
 35 40 45  
 Ala Glu Arg Ala Ala Val Met Gln Ala Gln Arg Thr Ala Ala Ala Ala  
 50 55 60  
 Ala Leu Glu Asn Ala Ala Leu Gln Ala Ala Glu Leu Pro Val Asp Ile  
 25 65 70 75 80  
 Glu Arg Arg Ile Arg Pro Ile Glu Gln Gln Val His His Ile Ala Asp  
 85 90 95  
 Ala Leu Glu Ala Leu Glu Thr Ala Ala Ala Ala Glu Glu Ala Asp  
 100 105 110  
 Ala Ala Arg Asp Ala Glu Arg Glu Gly Ala Ala Asp Gly Ala Ala Pro  
 30 115 120 125  
 Ser Pro Thr Ala Gly Pro Ala Ala Ala Glu Met Glu Val Gln Ile Val  
 130 135 140  
 Arg Asn Asp Pro Pro Leu Arg Tyr Asp Thr Asn Leu Pro Val Asp Leu  
 35 145 150 155 160  
 Leu His Met Val Tyr Ala Gly Arg Gly Ala Ala Gly Ser Ser Gly Val  
 165 170 175  
 Val Phe Gly Thr Trp Tyr Arg Thr Ile Gln Glu Arg Thr Ile Ala Asp  
 180 185 190  
 Phe Pro Leu Thr Thr Arg Ser Ala Asp Phe Arg Asp Gly Arg Met Ser  
 40 195 200 205  
 Lys Thr Phe Met Thr Ala Leu Val Leu Ser Leu Gln Ser Cys Gly Arg  
 210 215 220

[illegible]

40 (2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 91 amino acids

(B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

```

Val Val Asp Met Leu Ser Gly Ala Arg Gln Ala Ala Leu Val Arg Leu
10 1          5          10          15
Thr Ala Leu Glu Leu Ile Asn Arg Thr Arg Thr Asn Thr Thr Pro Val
          20          25          30
Gly Glu Ile Ile Asn Ala His Asp Ala Leu Gly Ile Gln Tyr Glu Gln
          35          40          45
15 Gly Leu Gly Leu Leu Ala Gln Gln Ala Arg Ile Gln Ala Lys Arg Phe
          50          55          60
Ala Thr Phe Asn Val Gly Ser Asp Tyr Asp Leu Leu Tyr Phe Leu Cys
65          70          75          80
Leu Gly Phe Ile Pro Gln Tyr Leu Ser Val Ala
20          85          90

```

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 444 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

```

Val Arg Val Pro Met Ala Ser Ala Glu Met Arg Glu Arg Leu Glu Ala
35 1          5          10          15
Pro Leu Pro Asp Arg Ala Val Pro Ile Tyr Val Ala Gly Phe Leu Ala
          20          25          30
Leu Tyr Asp Ser Gly Asp Pro Gly Glu Leu Ala Leu Asp Pro Asp Thr
          35          40          45
40 Val Arg Ala Ala Leu Pro Pro Glu Asn Pro Leu Pro Ile Asn Val Asp
          50          55          60
His Arg Ala Arg Cys Glu Val Gly Arg Val Leu Ala Val Val Asn Asp
65          70          75          80

```

Pro Arg Gly Pro Phe Phe Val Gly Leu Ile Ala Cys Val Gln Leu Glu  
                                     85                                    90                                    95  
 Arg Val Leu Glu Thr Ala Ala Ser Ala Ala Ile Phe Glu Arg Arg Gly  
                                     100                                    105                                    110  
 5 Pro Ala Leu Ser Arg Glu Glu Arg Leu Leu Tyr Leu Ile Thr Asn Tyr  
                                     115                                    120                                    125  
 Leu Pro Ser Val Ser Leu Ser Thr Lys Arg Arg Gly Asp Glu Val Pro  
                                     130                                    135                                    140  
 Pro Asp Arg Thr Leu Phe Ala His Val Cys Ala Ile Gly Arg Arg Leu  
 10 145                                    150                                    155                                    160  
 Gly Thr Ile Val Thr Tyr Asp Thr Ser Leu Asp Ala Ala Ile Ala Pro  
                                     165                                    170                                    175  
 Phe Arg His Leu Asp Pro Ala Thr Arg Glu Gly Val Arg Arg Glu Ala  
                                     180                                    185                                    190  
 15 Ala Glu Ala Glu Leu Ala Gly Arg Thr Trp Ala Pro Gly Val Glu Ala  
                                     195                                    200                                    205  
 Leu Thr His Thr Leu Leu Ser Thr Ala Val Asn Asn Met Met Leu Arg  
                                     210                                    215                                    220  
 Asp Arg Trp Ser Leu Val Ala Glu Arg Arg Arg Gln Ala Gly Ile Ala  
 20 225                                    230                                    235                                    240  
 Gly His Thr Tyr Leu Gln Ala Ser Glu Lys Phe Lys Ile Trp Gly Ala  
                                     245                                    250                                    255  
 Glu Ser Ala Pro Ala Pro Glu Arg Gly Tyr Lys Thr Gly Ala Pro Gly  
                                     260                                    265                                    270  
 25 Ala Met Asp Thr Ser Pro Ala Ala Ser Val Pro Ala Pro Gln Val Ala  
                                     275                                    280                                    285  
 Val Arg Ala Arg Gln Val Ala Ser Ser Ser Ser Ser Ser Ser Ser Phe  
                                     290                                    295                                    300  
 Pro Ala Pro Ala Asp Met Asn Pro Val Ser Ala Ser Gly Ala Pro Ala  
 30 305                                    310                                    315                                    320  
 Pro Pro Pro Pro Gly Asp Gly Ser Tyr Leu Trp Ile Pro Ala Phe His  
                                     325                                    330                                    335  
 Tyr Asn Gln Leu Val Thr Gly Gln Ser Ala Pro His His Pro Pro Leu  
                                     340                                    345                                    350  
 35 Thr Ala Cys Gly Leu Pro Ala Ala Gly Thr Val Ala Tyr Gly His Pro  
                                     355                                    360                                    365  
 Gly Ala Gly Pro Ser Pro His Tyr Pro Pro Pro Pro Ala His Pro Tyr  
                                     370                                    375                                    380  
 Pro Gly Tyr Ala Val Arg Gly Pro Gln Ser Pro Gly Gly Pro Asp Arg  
 40 385                                    390                                    395                                    400  
 Arg Ala Gly Gly Gly His Arg Arg Arg Pro Pro Gly Gly Trp Ala Ser  
                                     405                                    410                                    415  
 Gly Gly Arg Arg Arg Pro Arg Asp Pro Gly Val Gly Glu Pro Pro Pro

420 425 430  
 Thr Arg Gly Gly Ala Ala Gly Val Arg Leu Arg Pro  
 435 440

5 (2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 250 amino acids  
 (B) TYPE: amino acid  
 10 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Met Leu Phe Ala Gly Pro Ser Pro Leu Glu Ala Gln Ile Ala Ala Leu  
 1 5 10 15  
 Val Gly Ala Ile Ala Ala Asp Arg Gln Ala Gly Gly Leu Pro Ala Ala  
 20 20 25 30  
 Ala Gly Asp His Gly Ile Arg Gly Ser Ala Asn Arg Arg Arg His Glu  
 35 40 45  
 Val Glu Gln Pro Glu Tyr Asp Cys Gly Arg Asp Glu Pro Asp Arg Asp  
 50 55 60  
 25 Phe Pro Tyr Tyr Pro Gly Glu Ala Arg Pro Glu Pro Arg Pro Val Asp  
 65 70 75 80  
 Ser Arg Arg Ala Ala Arg Gln Ala Ser Gly Phe Thr Ile Thr Ala Leu  
 85 90 95  
 Val Gly Ala Val Thr Ser Leu Gln Gln Glu Leu Ala His Met Arg Ala  
 30 100 105 110  
 Arg Thr His Ala Pro Tyr Gly Pro Tyr Pro Pro Val Gly Pro Tyr His  
 115 120 125  
 His Pro His Ala Asp Thr Glu Thr Pro Ala Gln Pro Pro Arg Tyr Pro  
 130 135 140  
 35 Ala Glu Ala Val Tyr Leu Pro Pro Pro His Ile Ala Pro Pro Gly Pro  
 145 150 155 160  
 Pro Leu Ser Gly Ala Val Pro Pro Pro Ser Tyr Pro Pro Val Ala Val  
 165 170 175  
 Thr Pro Gly Pro Ala Pro Pro Leu His Gln Pro Ser Pro Ala His Ala  
 40 180 185 190  
 His Pro Pro Pro Pro Pro Gly Pro Thr Pro Pro Pro Ala Ala Ser  
 195 200 205  
 Leu Pro Gln Pro Glu Ala Pro Gly Ala Glu Ala Gly Ala Leu Val Asn  
 265



210 215 220  
 Ala Ser Ser Ala Ala His Val Asn Val Asp Thr Ala Arg Ala Ala Asp  
 225 230 235 240  
 Leu Phe Val Ser Gln Met Met Gly Ser Arg  
 5 245 250

## (2) INFORMATION FOR SEQ ID NO:55:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 1161 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

CCTGGCACCT TGTAGCAGTC CACCTATAGG ATACCCAGC ACTTTGGGAG GCCGAGGCAG 60  
 ACGGATCACA AGTTCAGGAG ATCGAAACCA TCCTGGCCAA CATGGTAAAA CCCCGTCTCT 120  
 20 ACTAAAAACG TAAAAATTAN CTGAGTGTGG TGGTGTGTGC CTGTACTCCC AGCTACTTGG 180  
 GAGGCTGAGG CAAAAAATTC ACTTGAACCT GGGAGGTGGA GGTTCGACTG AGCTGAGGTC 240  
 ATGCCACTGC ACTCCAGCCT AGCAACAGAG TGAGACTCCA TCTCAAAAAA ATAAATAAAT 300  
 AAATAAATAA ATAAATAAAG ACATATGGAG GCCTTACTCT GTGCCAAGCA CTATGATGGG 360  
 CACAGGGAAC AACACACGG GCTCCCTGAG CACCAGCGGT GAGCCAGGCA CCGTGCCTGG 420  
 25 AGACCAACGT CTGGCGTTTT GTATGCGGAC ATGATACCCG GCACTCTCCC CTATGGCTAA 480  
 TGAATCATCG AGCTTCACCA GAGAAACGCG AACAGACCCC CTTGTCCATG AATTTGCTAA 540  
 TTGACCTCCC CCAACTCAGA CATCAACCCT GCATTACCAT AAATTACTGG CTAAGAAACA 600  
 CCCGTTCTCA ACCTGCTGGC CTCAATGGGT. TACACGTCCC ACAAACCCCT TTCCCAAGGT 660  
 GAAATACCAA CCTCGAAAAC TCGGAAAATT CAAGTTCAAC AACCTCCGGG ATTGCGGGCT 720  
 30 TTACCAGCGA AGCCCTCTCC AAAAATTTGC TCGCTTAGAC ATCCAACCGC TTCTCCACTA 780  
 AAGACCCCGG CTTCTTCTCA CCTCGGCGTT CTCTTGCAAA AAATACGCGT CTGTTAATCC 840  
 GCGCCCCTCT TCCTCACACA CTTCTCCCTT GCCTACTCAT ACCTCATCTC TCCTATAACC 900  
 CTCTCGCGAA AGAGCCCTG TCTCTCACCT GTTTAATACC ACCCATGCGG GCGGGTTGTC 960  
 CCTTAAATGG ATATTCTGAA CCTCAACCCT CCCCCATTG TAATTTGGCG TGTTGGCCCC 1020  
 35 CTCACCTGCC CCCTCCTCTC CCAAAGTCCG GGAATATCCG TTCCCTGGCC ACCCTCCTTC 1080  
 TTTATGAAAC CCGGCTTACC CCCCCGGGA AATAGGCCGT TTGGCTTTTG TGGCGCGCCC 1140  
 TTCCACCTTC CCCTCTACAA 1161

## (2) INFORMATION FOR SEQ ID NO:56:

40

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids  
 (B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Val Lys Tyr Gln Pro Arg Lys Leu Gly Lys Phe Lys Phe Asn Asn Leu  
 1 5 10 15  
 10 Arg Asp Cys Gly Gln Arg Ser Pro Leu Gln Lys Phe Ala Arg Leu Asp  
 20 25 30  
 Ile Gln Pro Leu Leu His  
 35

15

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 524 base pairs

(B) TYPE: nucleic acid

20

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

25

GCAGGTCGAT CTAGAAGTCC CCAGGGGTCA GGGGTCTCAC TTGAGAAGGT AGTCTGTCCG 60  
 TTCTCCAATC TCAACCTCCG TGTGGGAGA TCCACTGCTC ACTTCAAAGC TGTGAGACAG 120  
 AGTTGTTTGC GTCTGCAGAG GTTTCAGCTG CTTTTTGTG TTGTAGTCGT CGTCGTTGTT 180  
 GTTGTGTTG TTGTTTAGCT GTGCCCTGTC CCCAGAGGTG GAGTCTACAG AGACAGGCAG 240  
 30 GACTCCTTGA GCTGCTGTGA GCTCCACCCA GTTGGAGCTT CCCAGCTGCT TTGTTTATCT 300  
 ACTTAAGCCT CAGTAATGGC GGGCGCCCCT CCCCCAACCT CGCTGCTGCC TTGCCCCCAG 360  
 ATCGCAGACT GCTGTGCTAA CAACGAGGGA GGCCCTGTGG GCATGGGACC CTCCTGGCCA 420  
 NGTGTGGGAT ATANTCTCCT GGTGTGCCCC GTTGGTAAAA TTTCTGGGTA AACCCCATAT 480  
 TGGGGGTTTG AATTCCCCAA ATTTCCCAGT TTGTTTGTG TCT 524

35

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 76 amino acids

40

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

```

5  Val Glu Leu Thr Ala Ala Gln Gly Val Leu Pro Val Ser Val Asp Ser
   1             5             10             15
   Thr Ser Gly Asp Arg Ala Gln Leu Asn Asn Asn Asn Asn Asn Asp
      20             25             30
10 Asp Asp Tyr Asn Asn Lys Lys Gln Leu Lys Pro Leu Gln Thr Gln Thr
   35             40             45
   Thr Leu Ser His Ser Phe Glu Val Ser Ser Gly Ser Pro Asn Thr Glu
   50             55             60
   Val Glu Ile Gly Glu Arg Thr Asp Tyr Leu Leu Lys
   65             70             75

```

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

```

20 (A) LENGTH: 773 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

```

30 GGAAGAGGGC GCTGTTGCCC GCGCTCCTTG CGCGGTGGCG GCGGGGGGCA GCGGAGGCA      60
   GCGCGGGCGT GCGGGGCCCTC CGGCGCCTTC CCCCCGCCCT CGCTCGGGGG GCTGTTTCGCC      120
   CACTCTGCGT CGTCGTTGCC GCGTAATCC GCGTCGTCGC TGTCGTCCGC CTGGGGCACC      180
30 AGCAGCCAGC GCCGCAGGAG CGATGACGCG GCCGCGCGC TCTCGACCGC GGTTCCCCGAG      240
   TCGTACGCAG GGACCATTTC GGAGTCTGCG GTTGGGAACG CGCCGGGGCG CGGCACGGTT      300
   GGACCGCCCG GCGCGGGCCG GCGCCGGGGA CCCC GGCGCG GGGGACTCCG GCGGGACATG      360
   GAGGGCGGCT GGGCTCGGCC TATGCCCGGA TCCGATCGC GTCTGGGCGG GAGATTTCAC      420
   TCGGCACGCA TGCACGTCTC CCCCCCCCCC CGTGGTTGCC TATGAACTA CCCCCTCCCG      480
35 CTGGTGTCG CATTTCTGTC CGCGTTGCCG GCCTTCTTTG CGGCGCGTGG CTTGACTGGG      540
   ATCCCCCTCC CTCTCCCTTC CCCTCCGGGA TTCACCCCG GGGGGGGTTT TTCTGGGGGG      600
   GGGGTAAATA GCTGTCTGTC CCCTCCCCAC CGTTTCCTCC CTGGACTCCA CGGCGCTCCA      660
   TAACTCTCTC CTGGTCCACC CCCCATTCCC CACATGGCCT TTGCTTTTCA ACCCCCCCTT      720
40 CCGGTTGGGC TGCATATCAA TTTCCTTCTC CCCC GGCGGGA TCCCCTATTA CG      773

```

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 121 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

10 Met Ser Arg Arg Ser Pro Arg Arg Arg Gly Pro Arg Arg Arg Pro Arg  
     1                    5                    10                    15  
     Pro Gly Gly Pro Thr Val Pro Arg Pro Gly Ala Phe Pro Thr Ala Asp  
                     20                    25                    30  
     Ser Gln Met Val Pro Ala Tyr Asp Ser Gly Thr Ala Val Glu Ser Ala  
 15                    35                    40                    45  
     Pro Ala Ala Ser Ser Leu Leu Arg Arg Trp Leu Leu Val Pro Gln Ala  
                     50                    55                    60  
     Asp Asp Ser Asp Asp Ala Asp Tyr Ala Gly Asn Asp Asp Ala Glu Trp  
     65                    70                    75                    80  
 20 Ala Asn Ser Pro Pro Ser Glu Gly Gly Gly Lys Ala Pro Glu Ala Pro  
                     85                    90                    95  
     His Ala Ala Pro Ala Ser Ala Cys Pro Pro Pro Pro Pro Arg Lys Glu  
                     100                    105                    110  
     Arg Gly Gln Gln Arg Pro Leu Pro Xaa  
 25                    115                    120

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 981 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

CCCCGCGGAG AATGAGCCCA CACCCAGATT GTTCACCGCC CCCCATTTCC CCCCCCCCGG 60  
 GTATACACCN AGGGAAGG TTTTCCCCC CCCCCCGGA TCAAATTTCC CCCACAAGAA 120  
 40 CCGAGTTCCA GGTAAAGTTT AGTTGGGTGC CCTTCCAGG TTGACGGGGG TCGCCAATGT 180  
 CCCAGCGGGG GTTGGCGCCC TCAGGGGGG NGGGGCCAGC CCCC CGGGC GGTGCGCCAC 240  
 CAACTTCCAA GCCGCGGCC GCGAGGCCA GCACGGTCCC CGGGGGGCG GTGGCAGACG 300  
 CCCAGCGTAT CTGCGGGGCG GGGCCGCGT CCGCGTCGTC GCGCAGCACC AGCGGGGGCG 360

```

CGTCGCCGTC GGGCTAGAGC AGCGCCCGCG CGCAGAACTC CCGCCGCGGC CCGAGCAGAT 420
CAGCCGGGGCC GCCGCGCACG GTGTCGCGCC CCAGCGCCAC GTAGACGGGC CGCAGCGGCG 480
CGCCCAGGCC CCAGCGCGCG CAGGCGCGGT GCGAGTGCGC CTCGTCCTCG CAGAAGTCCG 540
GCGCGCCGGG CGCCATGGCG TCGCCCGCGC CCGAGGCGGC GGCCCGGCCG TCCAGCGCCG 600
5 GGAGCACGGC GCGGCGGTAC TCGCGCGGGG ACATGGGCAC CAGCGTGTCT GGGCCGAAGC 660
GCGTGCGCAC GCGGTACCGC ACGTTGGCCC CGCGGCAGAG GCGCAGCGGC GGCGCGTCGG 720
GGTACATGCG CGCGTGCGCG GTCTCCACGC GCGCGAATAC CCCGGCCCTA ACACTCTGCC 780
GGATGCCATC ACGGTGCTGC GCTTGTTCCG CGCCCCCGGT CTTCGCACGG CGCTCTGTCT 840
TGGCGGGCTC CTCCTCCCTA GGTATTTTTT GGGTTCTTTC CTCTAAAAAA CCCGGGGCCT 900
10 CTTTGGGGG GGCCTTTTCC TCCCGGTCCC CTCCCCGGTT TGTGAACCAA CTAAATATAG 960
GCCGGTGGTT CCCCAGGCC 981

```

## (2) INFORMATION FOR SEQ ID NO:62:

## 15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 122 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

```

25 Val Glu Thr Ala His Ala Arg Met Tyr Pro Asp Ala Pro Pro Leu Arg
   1           5           10           15
Leu Cys Arg Gly Ala Asn Val Arg Tyr Arg Val Arg Thr Arg Phe Gly
   20           25           30
Pro Asp Thr Leu Val Pro Met Ser Pro Arg Glu Tyr Arg Arg Ala Val
30   35           40           45
Leu Pro Ala Leu Asp Gly Arg Ala Ala Ala Ser Gly Ala Gly Asp Ala
   50           55           60
Met Ala Pro Gly Ala Pro Asp Phe Cys Glu Asp Glu Ala His Ser His
   65           70           75           80
35 Arg Ala Cys Ala Arg Trp Gly Leu Gly Ala Pro Leu Arg Pro Val Tyr
   85           90           95
Val Gly Arg Asp Thr Val Arg Gly Gly Pro Ala Asp Leu Leu Gly Pro
   100          105          110
Arg Arg Glu Phe Cys Ala Arg Ala Leu Leu
40   115          120

```

## (2) INFORMATION FOR SEQ ID NO:63:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 644 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 5 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

```

10 GCATGCCTGC AGGTCGACTC TAAAGGATCC CCCAGCTGCC TCTCCNTGG AAGACATATA      60
   TCTTCTTGTC CCGGACCAGC TTCAGTGGGG CCAGTGGGCC CTCTAGGGCA ACGGTCACCT      120
   GTATCGTGGC ATCCGTGTGC ACTGGCCCCA TTTCCATTGA AAAAATCAGG GTGTCACGGG      180
   TTCTGAGGCT GCCATTGAGG CGGTAAAAAA TGGCCCCGTC TAGCAGGTCC TCCTGGGAGA      240
   AGGCTGTGAC TGA CTGAGTG GCCCAAAGCA ACTGCCCCCA GCGACGGCCG GCTGTGACAT      300
15 GGTAGTGCAC CTCATCCCCA CTGCGGATGT CAAGGTGGT GTCCANGTGG AGCTCANCTG      360
   TGTCAATGGT GCCCTGACCT CCTTGAGGGA CCACAAGGCC AAAACCGTTG GCCACACGGA      420
   TGTAGGGCTC CGAGGCCTGC ACCTCCANCA NCACAGTGGC CTGGTGTTGC CCATCGGACA      480
   CACCTGCAGC TGGATCCAGC CATGATCAGC CCAAGTGCAT GAACAGGACT CGCCTCTTTC      540
   TGAGGTCCTC CTGGGTGAAG CGGTAAATGG GCCACGTGGG CTCATCCACG GCCACGATAC      600
20 TGCCAAAAAA GAAGTCCTTG CGGGTCAGGG TACCGANCTC AAA                      644

```

## (2) INFORMATION FOR SEQ ID NO:64:

## (i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 151 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 30 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

```

Val Ser Asp Gly Gln His Gln Ala Thr Val Xaa Xaa Glu Val Gln Ala
35 1          5          10          15
   Ser Glu Pro Tyr Ile Arg Val Ala Asn Gly Phe Gly Leu Val Val Pro
          20          25          30
   Gln Gly Gly Gln Gly Thr Ile Asp Thr Xaa Glu Leu His Xaa Asp Thr
          35          40          45
40 Asn Leu Asp Ile Arg Ser Gly Asp Glu Val His Tyr His Val Thr Ala
          50          55          60
   Gly Arg Arg Trp Gly Gln Leu Leu Trp Ala Thr Gln Ser Val Thr Ala
   65          70          75          80

```

Phe Ser Gln Glu Asp Leu Leu Asp Gly Ala Ile Phe Tyr Arg Leu Asn  
                             85                            90                            95  
 Gly Ser Leu Arg Thr Arg Asp Thr Leu Ile Phe Ser Met Glu Met Gly  
                             100                            105                            110  
 5 Pro Val His Thr Asp Ala Thr Ile Gln Val Thr Val Glu Gly Pro Leu  
                             115                            120                            125  
 Ala Pro Leu Lys Leu Val Arg His Lys Lys Ile Tyr Val Phe Xaa Gly  
                             130                            135                            140  
 Arg Gly Ser Trp Gly Ile Leu  
 10 145                            150

## (2) INFORMATION FOR SEQ ID NO:65:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 52 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Val Glu Leu Xaa Cys Val Asn Gly Ala Leu Thr Ser Leu Arg Asp His  
 25 1                            5                            10                            15  
 Lys Ala Lys Thr Val Gly His Thr Asp Val Gly Leu Arg Gly Leu His  
                             20                            25                            30  
 Leu Xaa Xaa His Ser Gly Leu Val Leu Pro Ile Gly His Thr Cys Ser  
                             35                            40                            45  
 30 Trp Ile Gln Pro  
                             50

## (2) INFORMATION FOR SEQ ID NO:66:

## 35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 585 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

```

GCTCAATCCT CGAATTCAGA AAACAGTTGC CATTTATCCC TTCTTGTCAG ACTTCAGACG      60
GGTGATTGAG ATTGGTAATA CTAGCGAGGC TTACGACGAA CTTTCCGTT ATTTCAAGTT      120
TCACGACCCC TTCCATGAAA CAGAGGAGGA AATCATGGCG ACCCTTGCCT ATATCGATGT      180
CAAAAATCTT GCCCATCGTA TCCAAGGTGA GGTAAAAATG ATTACGGGCT TGGACAACAA      240
5  TGTTCGCTAT CCCATTACCC AGTTTGCGAT TTATAATCGT CTGACCTGCG ATAAAACCTA      300
TCGCATCATG CCTGAGTATG CTCACGAAGC CATGAATGTA TTTGTCAATG ACCAAGTCTA      360
CAACTGGCTC TGTGGAAGTG AGATTCCTTT TAAATATCTA AAATAAGGAG TCGACTCTAA      420
GCACAAAATC TTAAAAATTA CAAACACGCA TAGTATCAGG GGATTAAAAA AACTTGATAC      480
TATGCGTTTT ATCATGGACA TATATTATAA TGAAACAAGA ACAGGACAAA TCGATCCGGA      540
10 CAGTCCAATC GATTCTAAC AATGTTTAA AAGTAAATGT GTCT      585

```

## (2) INFORMATION FOR SEQ ID NO:67:

## (i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 60 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

```

Met Ile Thr Gly Leu Asp Asn Asn Val Cys Tyr Pro Ile Thr Gln Phe
25  1           5           10           15
Ala Ile Tyr Asn Arg Leu Thr Cys Asp Lys Tyr Ile Met Pro Glu Tyr
      20           25           30
Ala His Glu Ala Met Asn Val Phe Val Asn Asp Gln Val Tyr Asn Trp
      35           40           45
30  Leu Cys Gly Ser Glu Ile Pro Phe Lys Tyr Leu Lys
      50           55           60

```

## (2) INFORMATION FOR SEQ ID NO:68:

35

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1237 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:



ACTGTATCGA TAAGCTTGAT CCATGGCGGT GGCCGACACA GGGAGGGGCG TCTTCTCCGG 60  
 AGCGGACGTA GGCGAATCGG AGGACGCGTC CTCGGTCAAG GCCATGAGGC GCCGCCCGGT 120  
 TAGGGGGGCC CGAACGTCGG GGTCAACECC CTCGGGGTCT GTCCGCAGGG CGCGTCAAAA 180  
 CCGCGGGCGG GGTGGGAAGG GGGCGTACGG ACCGTCACTT AGGGCCCCGG GGGCCCAATG 240  
 5 GGGTGGCAGG ACCCCGACGT CTTCCGTGGG TCGTGCCATC CGAATAAACG TGGGGCCCGT 300  
 AATCCCCACC AGCAGGCTCT GGGGAGCAAA ACCGACGCGT GGTAGGTCGC TGGGGGCGGC 360  
 GGGCGTCTGT GGGGGCAAAC AGCGCTCCCG GAAACGCAGG CCACAAAACC CGGGGTGGG 420  
 GGGGGAATAC CATACCGGGG GCACCTATCG CCACGGGCGG CCCGCGGGGA CCGGGGGGAC 480  
 TCACGGGCCG CCCTCCGCAC GCGCCTCCTG TGGGGGGGCG GTGGGGTTTT CTGCCTATTC 540  
 10 CCTTCCTTTC CTCATCCTCT TCCTTCCCTA CTTCCCCCTT TCTCATTTCT CCTCTCTTCT 600  
 GTTCACCCCT TACTCCGCTT CACTTGCTCT CTCTCTATTC ATTCCGTCCT CTACTTTTCT 660  
 CGTCTTTTCC TCTCTCTCCC CTCATCTATC TTCTCTTCAT CTCTTTTCTT CTCGCTCCCT 720  
 CTCTTTCCCA TCTTCCGTTC TTCTCTCACT TTCATATCAT TCTTGCCCTA CCCCAGACT 780  
 CGCTATTCTC TCTTCTCGC CACCAATCCT GTGTGTGTTT CTCGTTTTCT TCACACCTCG 840  
 15 TTCTATAGCT CACCACATCA TGTGCTTTCT CGTATCTCCT ATCCTCCTTA TCCTTCTCTT 900  
 TTCTTTCTCT CTCAACCGCT CCCTTCTGTT CCACAGACAC TCTCTCTGCT CTCTCTCATT 960  
 CTTGCGCTCT TGTATTCACT TCTCATCATT CTTACACTTT TCTCTCTCAT TCGCACCCAT 1020  
 CTACCGCTAC GTCATTCACA CCGCGATTTT TTCTAGCTCT ACCTATTCCT CCTCGACTTC 1080  
 TCTGTGCGAC TATACTCCCC TCTTCTTTCT GTCCTACACG TCTGAGATCA CATTGATCTT 1140  
 20 CCCTCACCCC TCTGCTCCTG ACTATACCTT ATTCTATTTA TTTCTCGACC CTTCTTCCC 1200  
 ATTCCTATT CTGACTTCT CTGCACCTCT CCTCAC 1237

## (2) INFORMATION FOR SEQ ID NO:69:

## 25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

30

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

35 Met Ala Leu Thr Glu Asp Ala Ser Ser Asp Ser Pro Thr Ser Ala Pro  
 1 5 10 15  
 Glu Lys Thr Pro Leu Pro Val Ser Ala Thr Ala Met Asp Gln Ala Tyr  
 20 25 30  
 Arg Tyr Ser Xaa Xaa  
 40 35

## (2) INFORMATION FOR SEQ ID NO:70:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2057 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

5 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

10	GGCGACACGA GAAGAGAGCA AGGAGGGGAG AGCGACAGAG TAACTACACA TCGGAGCGCA	60
	TAATGAAGGA CAGTCATAAG ATGAGAGCAA GAAGAGATGA GCAGAGTCAG AAGATATAGC	120
	GACATAGAAG TCAGAGGCGA CACGGGGTAC GTGAGAATGT CGAACGCAGG ACGGGCGGAG	180
	GAGAGGAAGG GGGCCCCGAGC GCAGAACGAA ACAGGGGAAC ACAGGAGAAC GGAAGGAGCG	240
	AGCAAAGGGG AACAGACAGA AGGGCGCGCA NAGCGGAACG CCGCGAGAGC CGAGAGCCAC	300
15	AACACGGCCA ACGCGGCGAG AGCGACAGGG CAGACAGGGC AACGCGCCAG ACGGAAGCGG	360
	CGAGGGACGA AGCGNAGAGC NGGAGAGGCA AGGACAAAGC AAAGGAGGAG GAGAGCGCAC	420
	AGAAACGAAG AAGACGAGCG CGACAAGCCA CGGGAACGGA ACGAAACCGA GGCCACGGAG	480
	GGAGCAAAGC GAAGGGAGGC CTCCTTCCTC ATCATATTTC GCATCATCAT CGTCTTCATC	540
	CGCATATTCC TCTGCGGGCG GGGCTGGTGG GAGCGTCGCG TCCGCGTCCG GCGCTGGGGA	600
20	GAGACGAGAA ACCTCCCTCG GCGCCGCGC TGCTGCGCCG CGGGGGCCGA GGAAGTGTGC	660
	CAGGAAGACG CGCCACGCGG AGGGCGGCCC CGAGCCCGGG GCGCGCGACC CGGCGCCCGG	720
	CCTCACGCGC TACCTGCCCA TCGCGGGGGT CTCGAGCGTC GTGGCCCTGG CGCCTTACGT	780
	GAACAAGACG GTCACGGGGG ACTGCCTGCC CGTCTTGGAC ATGGAGACGG GCCACATAGG	840
	GGCCTACGTG GTCCTCGTGG ACCAGACGGG GAACGTGGCG GACCTGCTGC GGGCCGCGGC	900
25	CCCCGCGTGG AGCCGCCGCA CCCTGCTCCC CGAGCACGCG CGCAACTGCG TGAGGCCCCC	960
	CGACTACCCG ACGCCCCCGG CGTCGGAGTG GAACAGCCTC TGGATGACCC CGGTGGGCAA	1020
	CATGCTCTTT GACCAGGGCA CCCTGGTGGG CGCGCTGGAC TTCCACGGCC TCCGGTCGCG	1080
	CCACCCGTGG TCTCGGGAGC AGGGCGCGCC CGCGCCGGCC GGCGACGCCC CCGCGGGCCA	1140
	CGGGGAGTAG GGGGAGCTAA CACTCGGCTT GCTGCCCCGAA GGAAGCCGC CCCCCACCGG	1200
30	ACCACCGGCC GAGGCGCCTC GGGGGCATGG GGATGTGGGG GGGGGGGGAA AACNNGGATC	1260
	ATATCCGGAT TCGGGGTGGG ATTGGGGGGG GTATGTTTTT TGTTTNTTTT TGTTTTNTTT	1320
	TTTTTNTTTT GGTGTTGGTT TTTTGGTTTT TTGTTTTTTT TTNGGGGAT TTTTGTTTTT	1380
	TTTTTTTTTT TTNTTTTTTC GTTTTTTTTT TTGTGTTTTT NTTTGGTNTT TGGTTTGTTT	1440
	TGTGTTTTTT TTTTTTTTNT TTTTTTTTTT TTTTGGGNTT TTTGTTTTTT TTGTTTGTTT	1500
35	CTTTGTTTTT TTTTNTTTTG TTTCGTGTTT TTCTTTTTTT TTTCCTTCCT TTTCCCCCCG	1560
	CTTTCCCCCC CCTNCTCCCC CTCCTCTTCT CTTTCTCTNN TTTTCCTCTT CCCTTTTTCT	1620
	TCCCGTCTCC CCTCTGCGTT TCCCTCTCCC TTTTCTTCCC TTCCCGCTTC TCCGTCCCTC	1680
	CTCTTTTCCC TCCTTCTCTT TTCTTCCCCT GCTGCCTCCC TCTCTCCTCC TGTCTTTTTC	1740
	CCTCTTTTTT CCCTCCCTCT GCCCTTTCTT CCCTTCTCCT CTCCCTCCC CTCTTTTCTT	1800
40	TCCTTCCTCG CGTCGTTCCCT CCCTCTTCTC TCTCTCTCTC CTCTNGTCCC CCCCCTCTTT	1860
	CTCTTCCCCC CTCTTTTCTT CTCCTCTCGT CCTCCTTCCC CCTCATTTTA GCCTCATCCC	1920
	CTCCATCCTA TTACTCCTCT ATTCTCCTCT CTCTCCCTCT TCCATCCCTT CCGCTCCTCC	1980
	CATTATTCCT CTAAGCTTGC CCTCCTCCAC CTTCTCTCTA TCTCAAGTCC TCCTCCCTCT	2040

CACTATTCGG TTCCCT

2057

## (2) INFORMATION FOR SEQ ID NO:71:

## 5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 125 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

15 Val Ala Pro Tyr Val Asn Lys Thr Val Thr Gly Asp Cys Leu Pro Val  
     1                    5                    10                    15  
 Leu Asp Met Gly His Ile Gly Ala Tyr Val Val Leu Val Asp Gln Thr  
                     20                    25                    30  
 Gly Asn Val Ala Asp Leu Leu Arg Ala Ala Ala Pro Ala Trp Ser Arg  
 20                    35                    40                    45  
 Arg Thr Leu Leu Pro Glu His Ala Arg Asn Cys Val Arg Pro Pro Asp  
                     50                    55                    60  
 Tyr Pro Thr Pro Pro Ala Ser Glu Trp Asn Ser Leu Trp Met Thr Pro  
                     65                    70                    75                    80  
 25 Val Gly Asn Met Leu Phe Asp Gln Gly Thr Leu Val Gly Ala Leu Asp  
                     85                    90                    95  
 Phe His Gly Leu Arg Ser Arg His Pro Trp Ser Arg Glu Gln Gly Ala  
                     100                    105                    110  
 Pro Ala Pro Ala Gly Asp Ala Pro Ala Gly His Gly Glu  
 30                    115                    120                    125

## (2) INFORMATION FOR SEQ ID NO:72:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 1468 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

GCACCGCCCG GAAAGGGATC CCGGGGGGAA CCCC GCCCCC GAGAGGCGAC CGGGGCAGAA

60

CCCCCGGCAC GGTGAGAGGN GACCCCCGGT TATCAGGCCC CCCTTTTTC CCGACCACCC 120  
 AGGAGGGGGG TTGGGGGTGT TCGGGGCGT GGGGTTTGGG GGCGGGGACG CTTGACGGGG 180  
 CAGACCCCCG CCCCCTTAA GCGGTCGGGG GACCCCATG GGCCGTGCGC CGCCCCCGA 240  
 CCCTTTGGGG GGGGCGAGGG AGGCAGGGAG CCCTGAGCCC GAGAGCGGGG GACAGGGGGG 300  
 5 GGGAGACGAG GGGTAGGAAT CCAAAGGACG CAGACCACCT TTGGTTACGG ACCCCTTTCT 360  
 CCCCCCTTC CGAACAAAA GCAGCGGGCG GGGGGCCGGG GTGAGGGAGG GACACGGGGG 420  
 ACACGGCAGC GGGGTCCCGC CTCACGCCCC GCGCCCTCTA AATCCCCCCC CGTTTCTTTG 480  
 TCAAGCAGCC CGCCGCCCCG CACGCCTGGG GGATGCTCAA CGACATGCAG TGGCTCGCCA 540  
 GCAGCGACTC GGAGGAGGAG ACCGAGGTGG GAATCTCTGA CGACGACCTT CACCGCGACT 600  
 10 CCACCTCCGA GGCGGGCAGC ACGGACACGG AGATGTTTGA GGCGGGCCTG ATGGACGCGG 660  
 CCACGCCCCC GGCCCGGCCC CCGCCGAGC GCCAGGGCAG CCCACGCCC GCCGACGCGC 720  
 AGGGATCCTG CGGGGGTGGG CCCGTGGGTG AGGAGGAAGC GGAAGCGGGA GGGGGGGGCG 780  
 ACGTGTGTGC CGTGTGCACG GACGAGATCG CCCC GCCCCT GCGCTGCCAG AGTTTTCCT 840  
 GCCTGCACCC CTTCTGCATC CCGTGCATGA AGACCTGGAT TCCGTGCGC AACACGTGTC 900  
 15 CCCTGTGCAA CACCCCGGTG GCGTACCTGA TAGTGGGCGT GACCGCCAGC GGGTCGTTC 960  
 GCACCATCCC GATAGTGAAC GACCCCGGA CCCGCGTGA GGCCGAGGCG GCCGTGCGGT 1020  
 CCGGCACGGC CGTGGACTTT ATCTGGACGG GCAACCCGCG GACGCCCCG CGCTCCCTGT 1080  
 CGCTGGGGGG ACACACGGTC CGGCCCTGT CGCCACCCC CCCGTGGCCC GGCACGGACG 1140  
 ACGAGGACGA TGACCCGCCC GACGGTGAGG GCGGGCGGGG GTCTGGCACT GGGCGGGGT 1200  
 20 CCGGCACTGG GCGGGGTCT GGCCTGGGC GGGGGTCCGG CACTGGGCGG GGTCTGGCG 1260  
 GGGGTCAGGC ACTAACCGGG GGTCCCGTC TCTGTCTCCC TCTGCAACCG GAACTAATTT 1320  
 CCGCCCCC CCCTAATACC TCCCGGCCG GGGCTGCTGT GCCGGGGCCA CCCCTGGTAA 1380  
 CTCCACCCCC CTTTTFACCT AACCTGCGCC CCGGGCCCC CCGGGGACT AACTCACCC 1440  
 GTGGCCCCC CTTCTGGGC CGGGGT 1468

## (2) INFORMATION FOR SEQ ID NO:73:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 319 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Met Leu Asn Asp Met Gln Trp Leu Ala Ser Ser Asp Ser Glu Glu Glu  
 1 5 10 15  
 40 Thr Glu Val Gly Ile Ser Asp Asp Asp Leu His Arg Asp Ser Thr Ser  
 20 25 30  
 Glu Ala Gly Ser Thr Asp Thr Glu Met Phe Glu Ala Gly Leu Met Asp  
 35 40 45

Ala Ala Thr Pro Pro Ala Arg Pro Pro Ala Glu Arg Gln Gly Ser Pro  
 50 55 60  
 Thr Pro Ala Asp Ala Gln Gly Ser Cys Gly Gly Gly Pro Val Gly Glu  
 65 70 75 80  
 5 Glu Glu Ala Glu Ala Gly Gly Gly Gly Asp Val Cys Ala Val Cys Thr  
 85 90 95  
 Asp Glu Ile Ala Pro Pro Leu Arg Cys Gln Ser Phe Pro Cys Leu His  
 100 105 110  
 10 Pro Phe Cys Ile Pro Cys Met Lys Thr Trp Ile Pro Leu Arg Asn Thr  
 115 120 125  
 Cys Pro Leu Cys Asn Thr Pro Val Ala Tyr Leu Ile Val Gly Val Thr  
 130 135 140  
 Ala Ser Gly Ser Phe Ser Thr Ile Pro Ile Val Asn Asp Pro Arg Thr  
 145 150 155 160  
 15 Arg Val Glu Ala Glu Ala Ala Val Arg Ser Gly Thr Ala Val Asp Phe  
 165 170 175  
 Ile Trp Thr Gly Asn Pro Arg Thr Ala Pro Arg Ser Leu Ser Leu Gly  
 180 185 190  
 Gly His Thr Val Arg Ala Leu Ser Pro Thr Pro Pro Trp Pro Gly Thr  
 20 195 200 205  
 Asp Asp Glu Asp Asp Asp Pro Pro Asp Gly Glu Gly Gly Arg Gly Ser  
 210 215 220  
 Gly Thr Gly Arg Gly Ser Gly Thr Gly Arg Gly Ser Gly Thr Gly Arg  
 225 230 235 240  
 25 Gly Ser Gly Thr Gly Arg Gly Ser Gly Gly Gly Gln Ala Leu Thr Gly  
 245 250 255  
 Gly Ser Arg Leu Cys Leu Pro Leu Gln Pro Glu Leu Ile Ser Arg Pro  
 260 265 270  
 Pro Pro Asn Thr Ser Pro Pro Gly Ala Ala Val Pro Gly Pro Pro Leu  
 30 275 280 285  
 Val Thr Pro Pro Pro Leu Leu Pro Asn Leu Arg Pro Pro Ala Pro Pro  
 290 295 300  
 Gly Thr Thr Leu Thr Arg Gly Pro Pro Phe Leu Gly Arg Gly Phe  
 305 310 315  
 35

## (2) INFORMATION FOR SEQ ID NO:74:

## (i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 620 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

AAAACGCACG AGTATTGCAC GAATAACCAA CCAAACAACC ACTCAGACCA TGTGGATCCA 60  
 5 TACCCTTACT TGGCAAAATG GGGCATTAGC CGTGAGCAGT TTAAGCATGA TATTGAGAAC 120  
 GGCTTGACGA TTGAAACAGG CTGGCAGAAG AATGACACTG GCTACTGGTA CGTACACTCA 180  
 GACGGCTCTT ATCCAAAAGA CAAGTTTGAG AAAATCAATG GCACTTGGTA CTACTTTGAC 240  
 AGTTCAGGCT ATATGCTTGC AGACCGCTGG AGGAAGCACA CAGACGGCAA CTGGTACTGG 300  
 TTCGACAACT CAGGCGAAAT GGCTACAGGC TGGAAGAAAA TCGCTGATAA GTGGTACTAT 360  
 10 TTCAACGAAG AAGGTGCCAT GAAGACAGGC TGGGTCAAGT ACAAGGACAC TTGGTACTAC 420  
 TTAAACGCTA AAGAAGGCGC CATGGTATCA AATGCCTTTA TCCACTCAGC CGGACGGAAC 480  
 AGGCTGGTAC TACCTCAAAC CAGACCGAAC ACTGGCAGAC AAGCCAGAAT TCACAGTAGA 540  
 CCCAGATGCG TTGATTACGT TAAAATAATA ATGGAATGTC TTTCAAATCA AAACCCCGCA 600  
 TATTATTAGG TCTTGAAAA 620

## (2) INFORMATION FOR SEQ ID NO:75:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 116 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Met Leu Ala Asp Arg Trp Arg Lys His Thr Asp Gly Asn Trp Tyr Trp  
 1 5 10 15  
 30 Phe Asp Asn Ser Gly Glu Met Ala Thr Gly Trp Lys Lys Ile Ala Asp  
 20 25 30  
 Lys Trp Tyr Tyr Phe Asn Glu Glu Gly Ala Met Lys Thr Gly Trp Val  
 35 40 45  
 Lys Tyr Lys Asp Thr Trp Tyr Tyr Leu Asn Ala Lys Glu Gly Ala Met  
 35 50 55 60  
 Val Ser Asn Ala Phe Ile His Ser Ala Gly Arg Asn Arg Leu Val Leu  
 65 70 75 80  
 Pro Gln Trp Asn Thr Gly Arg Gln Ala Arg Ile His Ser Arg Pro Arg  
 85 90 95  
 40 Trp Leu Asp Tyr Val Lys Ile Ile Met Glu Cys Leu Ser Asn Gln Asn  
 100 105 110  
 Pro Ala Tyr Tyr  
 115

## (2) INFORMATION FOR SEQ ID NO:76:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 2695 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

GGGAGAGAAG AGAGAGAGAG AGAGAGAAGG AGTAGGAGAG CGAGGAGAGG AGAATAAGGA 60  
 GTGAATGGAA GCAGTAAGCT AGATAGGCAG AGAGAGAGAG AACGGAGAGT AGGAGTGGAA 120  
 15 GAAGTGAAG TTGAGAACGA CAAGGAGAGA GAAAGGAAGA AAAGTAGAGA GACTAGAGAA 180  
 TAGAAGAGGA GAACAGAGAG GTAGGAGAAA GAGGAAAAGA AGAGAGAGAG AAGGCAGCGA 240  
 GAAAGAGAGG AGCAGGCGGA CAAGGAGAAG AGGGAGGAAG AGGAAAAGAG GAAGAGAGAA 300  
 GAAGAATGGT GGAGAGAGAA GAGGAAAGAG CACCCGCGCC ACCGAGGATT GGGAGATGAA 360  
 TTAGGGGCCC CTAAGAGGAC CGAAGACCCG GCGGTAGATT ATTCGCCCCG GAGGGCAAGG 420  
 20 GAGGTGCGAC GCAAAGTAAA TACACCACCA GGGAGGAGGG AAATATGAAC GCCGGCGGAG 480  
 ACCCGGGGCC CTAGATACTG GTGAAGACGA TCAAAAGTAT GGATATGCCG GTCGCCACCA 540  
 GCTTTTGGC CCCGGACGGA ACGCCGCTGC AGTACGCGCT ATGCTTCCCG GCCGTACCCG 600  
 ACAAACTCGG CGCGCTGCTG ATGCGTCCCG AGGCGGCCTG CGTGGCGGCC CCGCTTCCGA 660  
 CGGACGTCCT CGAATCGGCC CCGACGCTCA CGGCCATGTA CGTGCTGACC GTCGTGAACC 720  
 25 GGCTCCAGCT GGCCCTCAGC GACGCCCAGG CCGCCAACTT TCAGCTGTTC GGTCGCTTCG 780  
 TGGCCATCG CCAGGCGACG TGGGGCGCCT CGATGGACGC GGCGGCCGAG CTGTACGTCG 840  
 CCCTCGTCGC CACCACCTC ACGCGCGAGT TTGGGTGTCG CTGGGCCCAG CTGGGTGGG 900  
 CGTCCGGAGC GGCGGCGCCG CGTCCGCGC CGGGCCCCCG GGGGTCCCAG CGCCACTGCG 960  
 TCGCCTTCAA CGAGAACGAC GTGCTGGTCG CGTGGTGGC CGGCGTTCCG GAACACATCT 1020  
 30 ACAACTTCTG CGCCTGGAC CTCGTTCCGC AGCACGAGTA CATGCACCTC ACCCTCGAAC 1080  
 GCGCGTTCGA GGACGCAGCG GAGTCCATGC TGTTCGTCCA GCGCCTGACC CCGCATCCCG 1140  
 ACGCCCGCAT CCGCGTGTG CCGACGTTT TGGACGGAGG CCCCCGACC CGGGGCCTCC 1200  
 TGTTCGGCAC GCGGCTGGCC GACTGGCGCC GGGGCAAGCT GTCCGAAACC GACCCGCTGG 1260  
 CGCCCTGGCG CTCGGCCTTG GAGCTCGGGA CCCAGCGCCG GGACGCCCCG GCGCTCGGGA 1320  
 35 AGCTCAGTCC GGCCAGGCC CTGGCGGCGG TGAGCGTCCT CGGGCGCATG TGTCTGCCGA 1380  
 GCGCCGCTTT GGCCGCGCTG TGGACCTGCA TGTTCCTCGA CGACTACACC GAGTACGACA 1440  
 GCTTCGACGC CCTCCTGGCC GCACGCTGG AGTCTGGCCA GACGCTCGGC CCGCGGGGG 1500  
 GGCGCGAGGC GTCCCTCCCC GAGGCCCCCC ACGCCCTCTA CCGACCCACG GGCCAGCACG 1560  
 TGGCCGTGCT GGCCGCCGCG ACCCACC GCA CCCCCGCGC GCGCGTTACG GCCATGGACC 1620  
 40 TGGTTCTGGC CGCGGTGCTC CTCGGCGCGC CCGTCGTGGT GGCGCTCCGC AACACCACGG 1680  
 CCTTCTCCCG CGAGTCGAA CTGGAAGTGT GCCTGACGCT CTTGACTCG CGCCCCGGCG 1740  
 GGCCGACGC CGCCCTGCGC GACGTCGTGT CGTCCGACAT CGAGACGTGG GCCGTCGGCC 1800  
 TCCTCCACAC CGATCTCAAC CCGATCGAAA ACGCGTGTCT GGCGGCGCAG CTCCCGCGCC 1860

5 TGTCGGCGCT CATCGCCGAG CGCCCTCTCG CCGACGGGCC CCCGTGCCTG GTCCTCGTGG 1920  
 ACATCTCCAT GACCCCGGTC GCGGTCTTGT GGAAGCCCC GGAGCCCCC GGCCCCCCTG 1980  
 ACGTGC GGTT TGTGGGCAGC GAGGCCACCG AGGAGCTTCC GTTTGTGGCT ACCGCGGGGG 2040  
 ACGTTCTTGC GGCGAGCGCC GCCGACGCGG ACCCCTTCTT CGCGCGGGCC ATCCTCGGGC 2100  
 5 GGCCCTTCGA CGCTCCCTC CTGACGGGGG AGCTGTTCCC GGGACACCCG GTTTACCAGC 2160  
 GCCCCCTCGC CGACGAGGCA GGTCCCTCTG CCCCAGCCGC CGCCCGCGAC CCGCGGGACC 2220  
 TTGCGGGGGG GGATGGCGGA TCGGGTCCCG AGGACCCCGC TGCCCCCCCC GCGCGGCAGG 2280  
 CGGACCCGGG GGTCTCGCC CCCACTTTCC TCACCGACGC CACCACCGGC GAGCCCGTCC 2340  
 CCCCTCGCAT GTGGGCCTGG ATCCACGGCC TGGAGGAGCT GGCGTCCGAG GACGCCGGCG 2400  
 10 GCCCCACGCC CAATCCGGCC CCGGCCTTAC TTCCCCCCCC CGCCACCGAT CAGTCCGTCC 2460  
 CCACGTCCCA GTACGCACCG CGGCCATCG GGCCGGCAGN TACGGCTCGC GAAACACGAC 2520  
 CGAGTGTCCC GCCTCAACAA AACACGGGGC GCGTGCCCGT GGCCCTCGG GANGACCCAC 2580  
 GGCCCTCGCC ACCACACCG AGTCCCCCG CGGATGCCGC GGTTCCTCCC CCGGCCTTTT 2640  
 CCGGGTTTGC CGCCGCTTTT TCCGCCGCCG TGCCGCGCGT GCGCAGATCC CGCC 2695  
 15

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 718 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

Val Lys Thr Ile Lys Ser Met Asp Met Pro Val Ala Thr Ser Phe Leu  
 1 5 10 15  
 30 Ala Pro Asp Gly Thr Pro Leu Gln Tyr Ala Leu Cys Phe Pro Ala Val  
 20 25 30  
 Thr Asp Lys Leu Gly Ala Leu Leu Met Arg Pro Glu Ala Ala Cys Val  
 35 40 45  
 Arg Pro Pro Leu Pro Thr Asp Val Leu Glu Ser Ala Pro Thr Val Thr  
 35 50 55 60  
 Ala Met Tyr Val Leu Thr Val Val Asn Arg Leu Gln Leu Ala Leu Ser  
 65 70 75 80  
 Asp Ala Gln Ala Ala Asn Phe Gln Leu Phe Gly Arg Phe Val Arg His  
 85 90 95  
 40 Arg Gln Ala Thr Trp Gly Ala Ser Met Asp Ala Ala Ala Glu Leu Tyr  
 100 105 110  
 Val Val Ala Thr Thr Leu Thr Arg Glu Phe Gly Cys Arg Trp Ala Gln  
 115 120 125



Leu Gly Trp Ala Ser Gly Ala Ala Ala Pro Arg Pro Pro Pro Gly Pro  
 130 135 140  
 Arg Gly Ser Gln Arg His Cys Val Ala Phe Asn Glu Asn Asp Val Leu  
 145 150 155 160  
 5 Val Val Ala Gly Val Pro Glu His Ile Tyr Asn Phe Trp Arg Leu Asp  
 165 170 175  
 Leu Val Arg Gln His Glu Tyr Met His Leu Thr Leu Glu Arg Ala Phe  
 180 185 190  
 Glu Asp Ala Ala Glu Ser Met Leu Phe Val Gln Arg Leu Thr Pro His  
 10 195 200 205  
 Pro Asp Ala Arg Ile Arg Val Leu Pro Thr Phe Leu Asp Gly Gly Pro  
 210 215 220  
 Pro Thr Arg Gly Leu Leu Phe Gly Thr Arg Leu Ala Asp Trp Arg Arg  
 225 230 235 240  
 15 Gly Lys Leu Ser Glu Thr Asp Pro Leu Ala Pro Trp Arg Ser Ala Leu  
 245 250 255  
 Glu Leu Gly Thr Gln Arg Arg Asp Ala Pro Ala Leu Gly Lys Leu Ser  
 260 265 270  
 Pro Ala Gln Ala Ala Val Ser Val Leu Gly Arg Met Cys Leu Pro Ser  
 20 275 280 285  
 Ala Ala Ala Leu Trp Thr Cys Met Phe Pro Asp Asp Tyr Thr Glu Tyr  
 290 295 300  
 Asp Ser Phe Asp Ala Leu Leu Ala Ala Arg Leu Glu Ser Gly Gln Thr  
 305 310 315 320  
 25 Leu Gly Pro Ala Gly Gly Arg Glu Ala Ser Leu Pro Glu Ala Pro His  
 325 330 335  
 Ala Leu Tyr Arg Pro Thr Gly Gln His Val Ala Val Leu Ala Ala Ala  
 340 345 350  
 Thr Thr Pro Ala Ala Arg Val Thr Ala Met Asp Leu Val Leu Ala Ala  
 30 355 360 365  
 Val Leu Leu Gly Ala Pro Val Val Val Arg Asn Thr Thr Ala Phe Ser  
 370 375 380  
 Arg Glu Ser Glu Leu Glu Leu Cys Leu Thr Leu Phe Asp Ser Arg Pro  
 385 390 395 400  
 35 Gly Gly Pro Asp Ala Ala Leu Arg Asp Val Val Ser Ser Asp Ile Glu  
 405 410 415  
 Thr Trp Ala Val Gly Leu Leu His Thr Asp Leu Asn Pro Ile Glu Asn  
 420 425 430  
 Ala Cys Leu Ala Ala Gln Leu Pro Arg Leu Ser Ala Leu Ile Ala Glu  
 40 435 440 445  
 Arg Pro Leu Ala Asp Gly Pro Pro Cys Leu Val Leu Val Asp Ile Ser  
 450 455 460  
 Met Thr Pro Val Ala Val Leu Trp Glu Ala Pro Glu Pro Pro Gly Pro

465                      470                      475                      480  
 Pro Asp Val Arg Phe Val Gly Ser Glu Ala Thr Glu Glu Leu Pro Phe  
                                  485                      490                      495  
 Val Ala Thr Ala Gly Asp Val Leu Ala Ala Ser Ala Ala Asp Ala Asp  
 5                      500                      505                      510  
 Pro Phe Phe Ala Arg Ala Ile Leu Gly Arg Pro Phe Asp Ala Ser Leu  
                                  515                      520                      525  
 Leu Thr Gly Glu Leu Phe Pro Gly His Pro Val Tyr Gln Arg Pro Leu  
                                  530                      535                      540  
 10 Ala Asp Glu Ala Gly Pro Ser Ala Pro Thr Ala Ala Arg Asp Pro Arg  
 545                      550                      555                      560  
 Asp Leu Ala Gly Gly Asp Gly Gly Ser Gly Pro Glu Asp Pro Ala Ala  
                                  565                      570                      575  
 Pro Pro Ala Arg Gln Ala Asp Pro Gly Val Leu Ala Pro Thr Phe Leu  
 15                      580                      585                      590  
 Thr Asp Ala Thr Thr Gly Glu Pro Val Pro Pro Arg Met Trp Ala Trp  
                                  595                      600                      605  
 Ile His Gly Leu Glu Glu Leu Ala Ser Glu Asp Ala Gly Gly Pro Thr  
                                  610                      615                      620  
 20 Pro Asn Pro Ala Pro Ala Leu Leu Pro Pro Pro Ala Thr Asp Gln Ser  
 625                      630                      635                      640  
 Val Pro Thr Ser Gln Tyr Ala Pro Arg Pro Ile Gly Pro Ala Xaa Thr  
                                  645                      650                      655  
 Ala Arg Glu Trp Ser Val Pro Pro Gln Gln Asn Thr Gly Arg Val Pro  
 25                      660                      665                      670  
 Val Ala Pro Arg Xaa Asp Pro Arg Pro Ser Pro Pro Thr Pro Ser Pro  
                                  675                      680                      685  
 Pro Ala Asp Ala Ala Val Pro Pro Pro Ala Phe Ser Gly Phe Ala Ala  
                                  690                      695                      700  
 30 Ala Phe Ser Ala Ala Val Pro Arg Val Arg Arg Ser Arg Arg  
 705                      710                      715

## (2) INFORMATION FOR SEQ ID NO:78:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2842 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

	GGGGGGAGGG	AAACAAGCCC	GATAGAGCGC	TAATGAGAAG	GGAGGATATA	ATGAGGGACA	60
	TTGGGGGAGG	GAAGGTAACA	GGAAGGTTTT	AGGAGCCCAA	GAAGACCGGA	GGACCCCCAA	120
	GACATTGGAG	GAAATGGCCG	AGGCCGTTAA	GGAGGGAAGA	GGTTCAGACA	TCCCGGCCCC	180
	CACCCGTACA	AGAGCCCAGG	AGCCACCAAG	CCCCGGGAGC	GGGAAAACAA	AAAGCCGCCC	240
5	AAGCGGCCCC	AGGCGACCCC	GCCCCCGAC	GCCAACGCGA	CCGTCGCCGC	CGGCCACGCC	300
	ACGCTGCGCG	CGCACCTGCG	GGAAATCAAG	GTCGAGAACG	CCGATGCCCA	GTTTTACGTG	360
	TGCCCCCCCC	CGACGGGCGC	CACGGTGGTG	CAGTTTGAGC	AGCCGCGCCG	CTGCCCCACG	420
	CGCCCCGAGG	GGCAGAACTA	CACGGAGGGC	ATCGCGGTGG	TCTTCAAGGA	GAACATCGCC	480
	CCGTACAAAT	TCAAGGCCAC	CATGTACTAC	AAAGACGTGA	CCGTGTCGCA	GGTGTGGTTC	540
10	GGCCACCGCT	ACTCCCAGTT	TATGGGGATA	TTCGAGGACC	GCGCCCCCGT	TCCCTTCGAG	600
	GAGGTGATCG	ACAAGATTAA	CGCCAAGGGG	GTCTGCCGCT	CCACGGCCAA	GTACGTGCGG	660
	AACAACATGG	AGACCACCGC	GTTTCACCGG	GACGACCACG	AGACCGACAT	GGAGCTCAAG	720
	CCGGCGAAGG	TCGCCACCGC	CACGAGCCGG	GGGTGGCACA	CCACCGACCT	CAAGTACAAC	780
	CCCTCGCGGG	TGGAGGCGTT	CCATCGGTAC	GGCACGACGG	TCAACTGCAT	CGTCGAGGAG	840
15	GTGGACGCGC	GGTCGGTGTA	CCCGTACGAT	GAGTTTGTGT	TGGCGACGGG	CGACTTTGTG	900
	TACATGTCCC	CGTTTTACGG	CTACCGGGAG	GGGTGCGACA	CCGAGCACAC	CAGCTACGCC	960
	GCCGACCGCT	TCAAGCAGGT	CGACGGCTTC	TACGCGCGCG	ACCTCACCAC	GAAGGCCCGG	1020
	GCCACGTGCG	CGACGACCCG	CAACTTGCTG	ACGACCCCCA	AGTTTACCGT	GGCCTGGGAC	1080
	TGGGTGCCGA	AGCGACCGGC	GGTCTGCACC	ATGACCAAGT	GGCAGGAGGT	GGACGAGATG	1140
20	CTCCGCGCCG	AGTACGGCGG	CTCCTTCCCG	TTCTCCTCCG	ACGCCATCTC	GACCACCTTC	1200
	ACCACCAACC	TGACCCAGTA	CTCGCTCTCG	CGCGTCGACC	TGGGCGACTG	CATTGGCCCG	1260
	GATGCCCCCG	AGGCCATCGA	CCGCATGTTT	GCGCGCAAGT	ACAACGCCAC	GCACATCAAG	1320
	GTGGGCCAGC	CGCAGTACTA	CCTGGCCACG	GGGGGCTTCC	TCATCGCGTA	CCAGCCCCTC	1380
	CTCAGCAACA	CGCTCGCCGA	GCTGTACGTG	CGGGAGTACA	TGCGGGAGCA	GGACCGCAAG	1440
25	CCCCGGAATG	CCACGCCCCG	GCCACTGCGG	GAGGCGCCCA	GCGCCAACGC	GTCCGTGGAG	1500
	CGCATCAAGA	CCACCTCCTC	GATCGAGTTC	GCCCGGCTGC	AGTTTACGTA	TAACCACATA	1560
	CAGCGCCACG	TGAACGACAT	GCTGGGGCGC	ATCGCCGTCG	CGTGGTGCGA	GCTGCAGAAC	1620
	CACGAGCTGA	CTCTCTGGAA	CGAGGCCCGC	AAGCTCAACC	CCAACGCCAT	CGCCTCCGCC	1680
	ACCGTCGGCC	GGCGGGTGAG	CGCGCGCATG	CTCGGAGACG	TCATGGCCGT	CTCCACGTGC	1740
30	GTGCCCGTCG	CCCCGGACAA	CGTGATCGTG	CAGAACTCGA	TGCGCGTCAG	CTCGCGGCCG	1800
	GGGACGTGCT	ACAGCCGCCC	CCTGGTCAGC	TTTCGGTACG	AAGACCAGGG	CCCCTGATC	1860
	GAGGGGCAGC	TGGGCGAGAA	CAACGAGCTG	CGCCTCACCC	GCGACGCGCT	CGAGCCGTGC	1920
	ACCGTGGGCC	ACCGGCGCTA	CTTCATCTTC	GGCGGGGGCT	ACGTGTACTT	CGAGGAGTAC	1980
	GCGTACTCTC	ACCAGCTGAG	TCGCGCCGAC	GTCACCACCG	TCAGCACCTT	CATCGACCTG	2040
35	AACATCACCA	TGCTGGAGGA	CCACGAGTTT	GTGCCCCTGG	AGGTCTACAC	GCGCCACGAG	2100
	ATCAAGGACA	GCGGCCGTGCT	GGACTACACG	GAGGTCCAGC	GCCGCAACCA	GCTGCACGAC	2160
	CTGCGCTTTG	CCGACATCGA	CACGGTCATC	CGCGCCGACG	CCAACGCCGC	CATGTTTCGCG	2220
	GGGCTGTGCG	CGTTCTTCGA	GGGGATGGGG	GACTTGGGGC	GCGCGGTCCG	CAAGGTAGTC	2280
	ATGGGAGTAG	TGGGGGGCGT	GGTGTGCGGC	GTCTCGGGCG	TGTCCTCCTT	TATGTCCAAC	2340
40	CCCTTCGGGG	CGCTTGCCGT	GGGGCTGCTG	GTCTGGCCCG	GCCTGGTTCG	GGCCTTCTTC	2400
	GCCTTCCGCT	ACGTCTTGCA	ACTGCAACGC	AATCCCATGA	AGGCCCTGTA	TCCGCTCACC	2460
	ACCAAGGAAC	TCAAGACTTC	CGACCCCGGG	GGCGTGGGCG	GGGAGGGGGA	GGAAGGCGCG	2520
	GAGGGGGGCG	GGTTTGACGA	GGCCAAGTTG	GCCGAGGCCC	GAGAAATGAT	CCGATATATG	2580

GNTTTGGTGT CGGCCATGGA GCGCACGGAA CACAAGGCCA GAAAGAAGGG CACGAGCGCC 2640  
 CTGCTCAGCT CCAAGGTCAC CAACATGGTT CTGCGCAAGC GCAACAAAGC CAGGTACTCT 2700  
 CCGCTCCACA ACGAGGACGA GGCCGGAGAC GAAGACGAGC TCTAAGGGAG GGGAGGGGAG 2760  
 CTGGGCTTGT GTATAAATAA AAAGACACCG ATGTTCAAAA ATACACATGA CTTCTNGGTA 2820  
 5 TGTNTGGGTA CCGAGCTCGA A 2842

## (2) INFORMATION FOR SEQ ID NO:79:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 787 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 15 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Val Cys Pro Pro Pro Thr Gly Ala Thr Val Val Gln Phe Glu Gln Pro  
 20 1 5 10 15  
 Arg Arg Cys Pro Trp Glu Gly Gln Asn Tyr Thr Glu Gly Ile Ala Val  
 20 25 30  
 Val Phe Lys Glu Asn Ile Ala Pro Tyr Lys Phe Lys Ala Thr Met Tyr  
 35 40 45  
 25 Tyr Lys Asp Val Thr Val Ser Gln Val Trp Phe Gly His Arg Tyr Ser  
 50 55 60  
 Gln Phe Met Gly Ile Phe Glu Asp Arg Ala Pro Val Pro Phe Glu Glu  
 65 70 75 80  
 Val Ile Asp Lys Ile Asn Ala Lys Gly Val Cys Arg Ser Thr Ala Lys  
 30 85 90 95  
 Tyr Val Arg Asn Asn Met Thr Ala Phe His Arg Asp Asp His Glu Thr  
 100 105 110  
 Asp Met Glu Leu Lys Pro Ala Lys Val Ala Thr Arg Thr Ser Arg Gly  
 115 120 125  
 35 Trp His Thr Thr Asp Leu Lys Tyr Asn Pro Ser Arg Val Glu Ala Phe  
 130 135 140  
 His Arg Tyr Gly Thr Thr Val Asn Cys Ile Val Glu Glu Val Asp Ala  
 145 150 155 160  
 Arg Ser Val Tyr Pro Tyr Asp Glu Phe Val Leu Ala Thr Gly Asp Phe  
 40 165 170 175  
 Val Tyr Met Ser Pro Phe Tyr Gly Tyr Arg Glu Gly Ser His Thr Glu  
 180 185 190  
 His Thr Ser Tyr Ala Ala Asp Arg Phe Lys Gln Val Asp Gly Phe Tyr  
 285

	195		200		205
	Ala Arg Asp Leu Thr Thr Lys Ala Arg Ala Thr Ser Pro Thr Thr Arg				
	210		215		220
	Asn Leu Leu Thr Thr Pro Lys Phe Thr Val Ala Trp Asp Trp Val Pro				
5	225		230		235 240
	Lys Arg Pro Ala Val Cys Thr Met Thr Lys Trp Gln Glu Val Asp Glu				
		245		250	255
	Met Leu Arg Ala Glu Tyr Gly Gly Ser Phe Arg Phe Ser Ser Asp Ala				
		260		265	270
10	Ile Ser Thr Thr Phe Thr Thr Asn Leu Thr Gln Tyr Ser Leu Ser Arg				
		275		280	285
	Val Asp Leu Gly Asp Cys Ile Gly Arg Asp Ala Arg Glu Ala Ile Asp				
		290		295	300
	Arg Met Phe Ala Arg Lys Tyr Asn Ala Thr His Ile Lys Val Gly Gln				
15	305		310		315 320
	Pro Gln Tyr Tyr Leu Ala Thr Gly Gly Phe Leu Ile Ala Tyr Gln Pro				
		325		330	335
	Leu Leu Ser Asn Thr Leu Ala Glu Leu Tyr Val Arg Glu Tyr Met Arg				
		340		345	350
20	Glu Gln Asp Arg Lys Pro Arg Asn Ala Thr Pro Ala Pro Leu Arg Glu				
		355		360	365
	Ala Pro Ser Ala Asn Ala Ser Val Glu Arg Ile Lys Thr Thr Ser Ser				
		370		375	380
	Ile Glu Phe Ala Arg Leu Gln Phe Thr Tyr Asn His Ile Gln Arg His				
25	385		390		395 400
	Val Asn Asp Met Leu Gly Arg Ile Ala Val Ala Trp Cys Glu Leu Gln				
		405		410	415
	Asn His Glu Leu Thr Leu Trp Asn Glu Ala Arg Lys Leu Asn Pro Asn				
		420		425	430
30	Ala Ile Ala Ser Ala Thr Val Gly Arg Arg Val Ser Ala Arg Met Leu				
		435		440	445
	Gly Asp Val Met Ala Val Ser Thr Cys Val Pro Val Ala Pro Asp Asn				
		450		455	460
	Val Ile Val Gln Asn Ser Met Arg Val Ser Ser Arg Pro Gly Thr Cys				
35	465		470		475 480
	Arg Pro Leu Val Ser Phe Arg Tyr Glu Asp Gln Gly Pro Leu Ile Glu				
		485		490	495
	Gly Gln Leu Gly Glu Asn Asn Glu Leu Arg Leu Thr Arg Asp Ala Leu				
		500		505	510
40	Glu Pro Cys Thr Val Gly His Arg Arg Tyr Phe Ile Phe Gly Gly Gly				
		515		520	525
	Tyr Val Tyr Phe Glu Glu Tyr Ala Tyr Ser His Gln Leu Ser Arg Ala				
		530		535	540

Asp Val Thr Thr Val Ser Thr Phe Ile Asp Leu Asn Ile Thr Met Leu  
 545 550 555 560  
 Glu Asp His Glu Phe Val Pro Leu Glu Val Tyr Thr Arg His Glu Ile  
 565 570 575  
 5 Lys Asp Ser Gly Leu Leu Asp Tyr Thr Glu Val Gln Arg Arg Asn Gln  
 580 585 590  
 Leu His Asp Leu Arg Phe Ala Asp Ile Asp Thr Val Ile Arg Ala Asp  
 595 600 605  
 Ala Asn Ala Ala Met Phe Ala Gly Leu Cys Ala Phe Phe Glu Gly Met  
 10 610 615 620  
 Gly Asp Leu Gly Arg Ala Val Gly Lys Val Val Met Gly Val Val Gly  
 625 630 635 640  
 Gly Val Val Ser Ala Val Ser Gly Val Ser Ser Phe Met Ser Asn Pro  
 645 650 655  
 15 Phe Gly Ala Val Gly Leu Leu Val Leu Ala Gly Leu Val Ala Ala Phe  
 660 665 670  
 Phe Ala Phe Arg Tyr Val Leu Gln Leu Gln Arg Asn Pro Met Lys Ala  
 675 680 685  
 Leu Tyr Pro Leu Thr Thr Lys Glu Leu Lys Thr Ser Asp Pro Gly Gly  
 20 690 695 700  
 Val Gly Gly Glu Gly Glu Glu Gly Ala Glu Gly Gly Gly Phe Asp Glu  
 705 710 715 720  
 Ala Lys Leu Ala Glu Ala Arg Glu Met Ile Arg Tyr Met Xaa Leu Val  
 725 730 735  
 25 Ser Ala Met Glu Arg Thr Glu His Lys Ala Arg Lys Lys Gly Thr Ser  
 740 745 750  
 Ala Leu Leu Ser Ser Lys Val Thr Asn Met Val Leu Arg Lys Arg Asn  
 755 760 765  
 Lys Ala Arg Tyr Ser Pro Leu His Asn Glu Asp Glu Ala Gly Asp Glu  
 30 770 775 780  
 Asp Glu Leu  
 785

## (2) INFORMATION FOR SEQ ID NO:80:

35

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4290 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

	GAGGAAGAGA	GGGGGGAGGG	GAAGAGAAAA	GAGAAAAGGA	AGAGGAGGGA	GGAGAAGAAG	60
	GGAAGGGTAA	GAGGGGAAGA	AGGGAANAAG	GAAAAAGGAG	TGTGAGAGGA	GGTAGGAGAT	120
	GAAGAGAAAA	GAGGGGGAGA	GAGAAGGGAA	AAAGAAGTGG	AAGGGAGGGA	GAAATAGGGG	180
5	AGAGGAGAAA	AGTAAGATTA	GGAGGTGGAG	AGGGGAGGAA	GAGGAATAAG	ATAGGTAGAG	240
	TAGGTGAAGG	TGGGAGAAGG	AGAGATGTAG	GGATAGGGAA	AAGGGGGGGG	AGGGGAGATG	300
	AGATAGAGAG	GAGGGGGGAA	AGGAAGGGGA	TGAAGAGGAG	GGGAGAGGGA	GGGGGGGAGA	360
	AAGGAGGTGA	GGGGGGAGGG	GGAAGAGGGG	GGAGGGGAGG	GGGGAGAGAA	GAANAGAAAA	420
	NNAAGNNNCC	CCGGCCCGCT	CCCCGNTGGA	GCCCCCTGGG	GACCCGACCC	TGTGGCGGGC	480
10	GCTGTATGCG	TGCGTCCTGG	CGGCCCTGGA	GCGCCAGACG	GGGCCGCTGG	CCCTCTTCGT	540
	CCCCTGCGC	CTGGGCTGGG	ACCCGCAGAC	GGGTCTGGTC	GTGAGGGTCG	AAAGGGCGTC	600
	GTGGGGCCCC	CCGGCCCGCT	CTCGCGCCGC	CCTCCTGGAC	GTGGAGGCCA	AGGTCAACTT	660
	CAACCCGCTG	GCCCTGSCCG	CGCGCGTCGC	CGAGCACCCC	GGCGCGCGGT	TGGCGTGGGC	720
	GCGCCTGGCC	GCCATTTCGA	ACAGCCCCCA	GTGCGCGTCC	TCCGCCCTGC	TCGCCGTCAC	780
15	CATCACGACG	AGGACCGCGC	GTTCGCGCG	CGAATACACC	ACCCTGGCGT	TTCCGCCGAC	840
	CAGCAAGGAG	GGCGCCTTCG	CGGACCTGGT	CGAGGTGTGC	GAGGTATGCC	TGCGGCCCCG	900
	CGGACACCCG	CATCGGTCA	CGGCGCGGGT	GCTGCTGCCG	CGCGGCTACA	ACTACTTCGT	960
	GAGCGCCGGC	GACGGGTCT	CCGCCCCGGC	GCTGGTCGCC	CTCTTCCGGC	AGTGGCATA	1020
	CACGGTCCAC	CCCGCCCCCG	GAGCCCTGGC	CCCCGTCTTC	GCTTTTCTGG	GGCCCGGGTT	1080
20	TGAGGTCCGG	GGAGGGCCCC	TCCAATACTT	TGCCGTGCTG	GGATTTCGG	GCTGGCCCCC	1140
	CTTTACCGTG	CCGGCCGCCG	CCGCCGCCGA	ATCGGTGCGT	GACCTGCTGC	GGGGCGCCGC	1200
	GTGCACCCAT	CCCCCTTGCC	CTGGGGGCCC	TGGCCCGCGG	TGGGCGCCTA	GGTCTTCCTG	1260
	CCCCCGCGGG	CATGGCCGGC	CGTGGCCTCG	GAGGCGGGCG	GCCGCCTCCT	GCCCCGCTTT	1320
	CGGGAAGCGG	TGGCGCGGTG	GCACCCACG	GCCACCACCA	TCCAACACT	CGACCCCCCG	1380
25	GCGGCCGTCG	GGCCGGTCTG	GACGGCGCGG	TTTTGTTTCT	CCGGGCTCCA	GGCCAGCTC	1440
	CTGGCCGCCC	TCGCGGCGCT	CGGGGAGGCC	GGGCTGCCGG	AAGCCCGGGG	GCGGGCGGGC	1500
	CTGGAAGGC	TGGACGCGCT	GGTGGCGGCC	GCCCCCTCGG	AGCCCTGGGC	CCGGGCCGTG	1560
	CTGGAGCGCC	TGGTGCCGGA	CGCGTGCGAC	GCCTGCCCCG	CGCTCCGGCA	GCTGCTCGGC	1620
	GGGGTCATGG	CCGCCGTCTG	CCTGCAGATC	GAGCAGACGG	CCAGCTCGGT	GAAGTTTGCG	1680
30	GTCTGCGGCG	GCACCGGGGC	TGCGTTCTGG	GGGCTGTTCA	ACGTGGACCC	CGGGGACGCG	1740
	GACGCCGCGC	ACGGCGCGAT	CCATGACGCC	CGCCGGGCCC	TCGAGGCGTC	CGTGCGCGCC	1800
	GTACTTTTCG	CCAACGGCAT	ACGCCCGCGC	CTCGCCCCCT	CCCTGGCGCT	AGAGGGCGTC	1860
	TACACCCACG	TCGTACCTG	GAGCCAGACC	GGGGCGTGGT	TCTGGAACTC	CCCGATGAC	1920
	ACCGACTTCC	TGCAGGGATT	TCCTCTCCGC	GGGCCCGCGT	ACGCCCGGGC	GGCCGAGGTT	1980
35	ATGCGCGACG	CGCTGAGACG	AATCCTCCGG	CGGCCGGCCG	CCGGCCCGCC	GGAGGAGGCC	2040
	GTGTGCGCGG	CCCGGGGCAT	CATGGAGGAC	GCCTGTGACC	GCTTTGTCTT	GGATGCCTTC	2100
	GGGAGGCGTC	TGGACGCGGA	GTACTGGAGC	GTTCTGACCC	CCCCGGGCGA	GGCCGACGAC	2160
	CCCCTGCCCC	AAACGGCCTT	CCGCGGAGGC	GCCCTGCTGG	ACGCGGAGCA	ATACTGGAGA	2220
	CGCGTCGTGC	GCGTATGTCC	CGGGGGCGGG	GAGTCGGTCG	GCGTCCCCGT	GGATCTGTAC	2280
40	CCGCGGCCCT	TGGTGCTCCC	CCCCGTGGAC	TGCGCCCATC	ACCTGCGCGA	GATCCTGCGC	2340
	GAGATTCAAC	TGGTGTTTAC	GGGGGTTCG	GAAGGCGTGT	GGGGCGAGGG	CGGGAGCTTT	2400
	GTGTACCCCT	TCGAGGAAAA	GATGCGGTTT	CTGTTTCCCT	GAATTGGGTC	AATAAACTGG	2460
	GGCCCCGTGC	TCCAACCTAC	CCCCGCGTGT	GCGCGCGTCC	GTATTTACTG	ACACGCGCCG	2520

	GTGTGGTTT TCTTCTATTT CTTTGTTCCT TCTATCATGT CTTTCCACCA CCAGCACCAC	2580
	CACCCCCCA CTTTCTCTCT TCGGTGCACA AGACACACAC ACAGGCCAC CACCATCCCC	2640
	CGAGAGATGA CACGACAGGT AGGGAACGTT CCATAAAAAA CACGTTTATT TTCCGGAGTT	2700
	AGCAAAACCG ATAGAAAAGC GACGAGGTCC GTCGTTTGGG GCTCCCCGAA AGCCACCAAT	2760
5	ACACCAGAGC CGAACGCAGG TCCTTGGATT TCCAGCAGCT TCCCATGACG CCGGCCGGGT	2820
	TATAGGCCAC ACAGTCCGTG CGGGGACGG GCCCGGGCAA CTGCAACGCA AAGCTCTGCG	2880
	GGGTGGCGCA AAACAGGGCC GAAAGGACGG GGGGCGGATT GTTGCCAGC AGGTAGTGGG	2940
	CCATGTACCA GTGGGGCAGG ACTAGCTCGT CGTCGAAGGG CTTACGCCC GCCATGCACA	3000
	TTAGCGTGT CAGGATCCAG TGGCAGCTGC GGAGGAGAAG GCAGCGGACG CGCTCGAAGG	3060
10	GAGACGCGTG GCGGGCCACA ACCCCCAACC ACGCCGACAC CTTGACAAAC AGCGAGGGCG	3120
	TGGCGTGGCT CCGGGGACAG TTCTCCAGGT ACATCAGCAG GCAGACGAGC TCGAAGTCCC	3180
	GGAGGTCCGT GGGGCGAAAC AAGGAGAGCC GGTGCAGCAG AACGAGGGGG GGCACGGCCA	3240
	GGCTGTGCAC GTGTCCTCG TTGGCCGTGA GGACGACGAC CGCAAAACCG CGATACGTCG	3300
	GCTGGTCCCC GCAGCGCTTA AAGTAATCCT CCGACGCCAC GTACGCGTCG TCGGAATCC	3360
15	CGTCCAGAAC GAACGCAAC CGCCCCCTGG GGGTGACGTC AACGCGCAGG ACGCTGGTCG	3420
	CGGTAAACCG CGGCTGGCGA TCGCTGACCT GCGCACCTC GCAGGCCATG CGCAGCAGCG	3480
	CCTGGTTGCT GATCCCCCTC GCCACCTCGA CCAGACTGCG GTCCCCGGCG ATGGCCTGTT	3540
	TGAGGATGGC GCGGGCCGTT CCCTCATCGG CGGGCGTGGG GTCGGCCATC CTTGCGTTGG	3600
	ACGCCCCAGC CCTGGTCCGG CGCACCCCTC GCGGTTCTCC CGGGCGACCG GGATCGGGTC	3660
20	CGGGTCCGGG ACCGGGACCC GCCCCGCGGG GACGCGCTCG CCCGAAATC GGCGGGGGTT	3720
	GGGGAGGGGG GCCGGGGCAG AGCCGCGTGC TGTACGTCCG CCACGAACAG GGCCGCGACG	3780
	TCTGTCAGGT ACGTCTGCAG GCGGGTTTTT TTAAAGACCG CCTCCCATAA CTCCTCCTCC	3840
	CCTAGGATGA CATCGGAGCC GGTGATGAGC GCGCCGCTC GGGGGGCGG AAGCACGTAC	3900
	TCGAAATACG GGGCCACGAA GGAGGCGATC GCGCCGCTAG AGTACGAGAT CGACGTTTCC	3960
25	TGGCCCTGGT TGTTCGGTG GCGCAGAATC TTGAAGCAGC GCACCAGCTC GTGCTCCCAG	4020
	AGGCGCGACA GCGCTCGAG GTCCTGGCCG TACGCGGGA TGTACTGGTG CTGGAACTG	4080
	TTGGCCACGT ACGTGTGTC GTCCATGGAC TTGCTGACGT CGATAATGTC GTAGTCGGCC	4140
	CGGAGAAGAT CCGCCTCCGC CGGGCGGGCC GCGCCTCCCC CGGCCGCCCC GTCCGCCGCG	4200
	CGATGCTCCC GCTCCAGCGC CCCCCTCGG GCGCGCCGA GCTCGCGGTC GCGCGCCTGC	4260
30	AGCTGGGTCG CCGGGGACAT CTAGAGTCG	4290

## (2) INFORMATION FOR SEQ ID NO:81:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 373 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 40 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:



Val Phe Val Pro Leu Arg Leu Gly Trp Asp Pro Gln Thr Gly Leu Val  
 1 5 10 15  
 Val Arg Val Glu Arg Ala Ser Trp Gly Pro Pro Ala Ala Pro Arg Ala  
 20 25 30  
 5 Ala Leu Leu Asp Val Glu Ala Lys Val Asn Phe Asn Pro Leu Ala Ala  
 35 40 45  
 Arg Val Ala Glu His Pro Gly Ala Arg Leu Ala Trp Ala Arg Leu Ala  
 50 55 60  
 Ala Ile Arg Asn Ser Pro Gln Cys Ala Ser Ser Ala Ser Leu Ala Val  
 10 65 70 75 80  
 Thr Ile Thr Thr Arg Thr Ala Arg Phe Ala Arg Glu Tyr Thr Thr Leu  
 85 90 95  
 Ala Phe Pro Pro Thr Ser Lys Glu Gly Ala Phe Ala Asp Leu Val Glu  
 100 105 110  
 15 Val Cys Glu Val Cys Leu Arg Pro Arg Gly His Pro His Arg Val Thr  
 115 120 125  
 Ala Arg Val Leu Leu Pro Arg Gly Tyr Asn Tyr Phe Val Ser Ala Gly  
 130 135 140  
 Asp Gly Phe Ser Ala Pro Ala Leu Val Phe Arg Gln Trp His Thr Thr  
 20 145 150 155 160  
 Val His Pro Ala Pro Gly Ala Pro Val Phe Ala Phe Leu Gly Pro Gly  
 165 170 175  
 Phe Glu Val Arg Gly Gly Pro Leu Gln Tyr Phe Ala Val Leu Gly Phe  
 180 185 190  
 25 Pro Gly Trp Pro Pro Phe Thr Val Pro Ala Ala Ala Ala Ala Glu Ser  
 195 200 205  
 Val Arg Asp Leu Leu Arg Gly Ala Ala Cys Thr His Pro Leu Cys Pro  
 210 215 220  
 Gly Gly Pro Gly Pro Arg Trp Ala Pro Arg Ser Ser Cys Pro Arg Gly  
 30 225 230 235 240  
 His Gly Arg Pro Trp Pro Arg Arg Arg Pro Ala Ala Ser Cys Pro Pro  
 245 250 255  
 Phe Gly Lys Arg Trp Arg Gly Gly Thr Pro Arg Pro Pro Pro Ser Asn  
 260 265 270  
 35 Tyr Ser Thr Pro Arg Arg Pro Ser Gly Arg Ser Gly Arg Arg Gly Phe  
 275 280 285  
 Val Ser Pro Gly Ser Arg Pro Ser Ser Trp Pro Pro Ser Arg Ala Ser  
 290 295 300  
 Gly Arg Pro Gly Cys Arg Lys Pro Gly Gly Gly Arg Ala Trp Lys Gly  
 40 305 310 315 320  
 Trp Thr Arg Trp Trp Arg Pro Pro Pro Arg Ser Pro Gly Pro Gly Pro  
 325 330 335  
 Cys Trp Ser Ala Trp Cys Arg Thr Arg Ala Thr Pro Ala Pro Arg Ser

340                      345                      350  
 Gly Ser Cys Ser Ala Gly Ser Trp Pro Pro Ser Ala Cys Arg Ser Ser  
 355                      360                      365  
 Arg Arg Pro Ala Arg  
 5                      370

## (2) INFORMATION FOR SEQ ID NO:82:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 380 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 15 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Val Ala Ser Glu Ala Ala Gly Arg Leu Leu Pro Ala Phe Arg Glu Ala  
 20                      5                      10                      15  
 Val Ala Arg Trp His Pro Thr Ala Thr Thr Ile Gln Leu Leu Asp Pro  
 20                      25                      30  
 Pro Ala Ala Val Gly Pro Val Trp Thr Ala Arg Phe Cys Phe Ser Gly  
 35                      40                      45  
 25 Leu Gln Ala Gln Leu Leu Ala Ala Gly Leu Gly Glu Ala Gly Leu Pro  
 50                      55                      60  
 Glu Arg Arg Ala Gly Leu Glu Arg Leu Asp Ala Leu Val Ala Ala Ala  
 65                      70                      75                      80  
 Pro Ser Glu Pro Trp Ala Arg Ala Val Leu Glu Arg Leu Val Pro Asp  
 30                      85                      90                      95  
 Ala Cys Asp Ala Cys Pro Ala Leu Arg Gln Leu Leu Gly Gly Val Met  
 100                      105                      110  
 Ala Ala Val Cys Leu Gln Ile Glu Gln Thr Ala Ser Ser Val Lys Phe  
 115                      120                      125  
 35 Ala Val Cys Gly Gly Thr Gly Ala Ala Phe Trp Gly Leu Phe Asn Val  
 130                      135                      140  
 Asp Pro Gly Asp Ala Asp Ala Ala His Gly Ala Ile His Asp Ala Arg  
 145                      150                      155                      160  
 Arg Ala Leu Glu Ala Ser Val Arg Ala Val Leu Ser Ala Asn Gly Ile  
 40                      165                      170                      175  
 Arg Pro Arg Leu Ala Pro Ser Leu Ala Leu Glu Gly Val Tyr Thr His  
 180                      185                      190  
 Val Val Thr Trp Ser Gln Thr Gly Ala Trp Phe Trp Asn Ser Arg Asp

```

      195              200              205
Asp Thr Asp Phe Leu Gln Gly Phe Pro Leu Arg Gly Pro Ala Tyr Ala
      210              215              220
Ala Ala Ala Glu Val Met Arg Asp Ala Leu Arg Arg Ile Leu Arg Arg
5  225              230              235              240
Pro Ala Ala Gly Pro Pro Glu Glu Ala Val Cys Ala Arg Ile Met Glu
      245              250              255
Asp Ala Cys Asp Arg Phe Val Leu Asp Ala Phe Gly Arg Arg Leu Asp
      260              265              270
10 Ala Glu Tyr Trp Ser Val Leu Thr Pro Pro Gly Glu Ala Asp Asp Pro
      275              280              285
Leu Pro Gln Thr Ala Phe Arg Gly Gly Ala Leu Leu Asp Ala Glu Gln
      290              295              300
Tyr Trp Arg Arg Val Val Arg Val Cys Pro Gly Gly Gly Glu Ser Val
15 305              310              315              320
Gly Val Pro Val Asp Leu Tyr Pro Arg Pro Leu Val Leu Pro Pro Val
      325              330              335
Asp Cys Ala His His Leu Arg Glu Ile Leu Arg Glu Ile Gln Leu Val
      340              345              350
20 Phe Thr Gly Val Leu Glu Gly Val Trp Gly Glu Gly Gly Ser Phe Val
      355              360              365
Tyr Pro Phe Glu Glu Lys Met Arg Phe Leu Phe Pro
      370              375              380

```

25 (2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 302 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

```

Val Arg Arg Thr Arg Ala Gly Asn Ala Gly Met Ala Asp Pro Thr Pro
1          5          10          15
Ala Asp Glu Gly Thr Ala Ala Ala Ile Leu Lys Gln Ala Ile Ala Gly
40          20          25          30
Asp Arg Ser Leu Val Glu Val Ala Glu Gly Ile Ser Asn Gln Ala Leu
      35          40          45
Leu Arg Met Ala Cys Glu Val Arg Gln Val Ser Asp Arg Gln Pro Arg
292

```

	50		55		60	
	Phe Thr Ala Thr Ser Val	Leu Arg Val Asp Val Thr Pro Arg Gly Arg				
	65	70	75	80		
	Leu Arg Phe Val Leu Asp Gly Ser Ser Asp Asp Ala Tyr Val Ala Ser					
5	85	90	95			
	Glu Asp Tyr Phe Lys Arg Cys Gly Asp Gln Pro Tyr Gly Phe Ala Val					
	100	105	110			
	Val Val Leu Thr Ala Asn Glu Asp His Val His Ser Leu Ala Val Pro					
	115	120	125			
10	Pro Leu Val Leu Leu His Arg Leu Ser Leu Phe Arg Pro Thr Asp Leu					
	130	135	140			
	Arg Asp Phe Glu Leu Val Cys Leu Leu Met Tyr Leu Glu Asn Cys Pro					
	145	150	155	160		
	Arg Ser His Ala Thr Pro Ser Leu Phe Val Lys Val Ser Ala Trp Leu					
15	165	170	175			
	Gly Val Val Ala Arg His Asp Phe Glu Arg Val Arg Cys Leu Leu Leu					
	180	185	190			
	Arg Ser Cys His Trp Ile Leu Asn Thr Leu Met Cys Met Ala Gly Val					
	195	200	205			
20	Lys Pro Phe Asp Asp Glu Leu Val Leu Pro His Trp Tyr Met Ala His					
	210	215	220			
	Tyr Leu Leu Ala Asn Asn Pro Pro Pro Val Leu Ser Ala Leu Phe Cys					
	225	230	235	240		
	Ala Thr Pro Gln Ser Phe Ala Leu Gln Leu Pro Gly Pro Val Pro Arg					
25	245	250	255			
	Thr Asp Cys Val Ala Tyr Asn Pro Ala Gly Val Met Gly Ser Cys Trp					
	260	265	270			
	Lys Ser Lys Asp Leu Arg Ser Ala Leu Val Tyr Trp Trp Leu Ser Gly					
	275	280	285			
30	Ser Pro Lys Arg Arg Thr Ser Ser Leu Phe Tyr Arg Phe Cys					
	290	295	300			

## (2) INFORMATION FOR SEQ ID NO:84:

## 35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 236 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

40

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Met Ser Pro Ala Thr Gln Leu Gln Ala Arg Asp Arg Glu Leu Arg Arg  
 1 5 10 15  
 Ala Gln Ala Gly Ala Leu Glu Arg Glu His Arg Ala Ala Asp Arg Ala  
 5 20 25 30  
 Ala Gly Gly Gly Ala Gly Arg Pro Ala Glu Ala Asp Leu Leu Arg Ala  
 35 40 45  
 Asp Tyr Asp Ile Ile Asp Val Ser Lys Ser Met Asp Asp Asp Thr Tyr  
 50 55 60  
 10 Val Ala Asn Ser Phe Gln His Gln Tyr Ile Pro Ala Tyr Gly Gln Asp  
 65 70 75 80  
 Leu Glu Arg Leu Ser Arg Leu Trp Glu His Glu Leu Val Arg Cys Phe  
 85 90 95  
 Lys Ile Leu Arg His Arg Asn Asn Gln Gly Gln Glu Thr Ser Ile Ser  
 15 100 105 110  
 Tyr Ser Ser Gly Ala Ile Ala Ser Phe Val Ala Pro Tyr Phe Glu Tyr  
 115 120 125  
 Val Leu Arg Ala Pro Arg Ala Gly Ala Leu Ile Thr Gly Ser Asp Val  
 130 135 140  
 20 Ile Leu Gly Glu Glu Glu Leu Trp Glu Ala Val Phe Lys Lys Thr Arg  
 145 150 155 160  
 Leu Gln Thr Tyr Leu Thr Asp Val Ala Ala Leu Phe Val Ala Asp Val  
 165 170 175  
 Gln His Ala Ala Leu Pro Arg Pro Pro Ser Pro Thr Pro Ala Asp Phe  
 25 180 185 190  
 Arg Ala Ser Asp Arg Gly Gly Ser Arg Ser Arg Thr Arg Thr Arg Ser  
 195 200 205  
 Arg Ser Pro Gly Arg Thr Pro Arg Gly Ala Pro Asp Gln Gly Trp Gly  
 210 215 220  
 30 Val Gln Arg Arg Asp Gly Arg Pro His Ala Arg Arg  
 225 230 235

## (2) INFORMATION FOR SEQ ID NO:85:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3664 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

	GTGTGTTTTT	GTTTGTCTCC	ACTTGAAGAG	GCGTAAATTT	GAGTTTCTAG	GGGGGGGCCA	60
	GAGGAAAACC	ANCAAGGCC	TTTAGGTTTT	CGCCCTTNGN	GGCCCNCTGGT	AACGTTTTAT	120
	CCGGNGTTAT	CCAGATAGGG	GAGACCCANA	CCCCCTTGAG	GAGGNAAACC	TTTCCCCACC	180
	CGNACCCGCG	CCCCAGATTT	AGTGAGGGGG	ANGGAAGAGC	CCCAAACACC	NACCCCTTTT	240
5	CCGGNGGGTN	GCGATTTAAT	ATGCANTGCA	GACAGTTCTC	GATTGGAACG	GGCATGGCGC	300
	AACCANTNAT	GGGTNGTCAT	CTTGACACCC	GCTACATTAA	GTTCGTTTGA	AGTGGGGATG	360
	GGGGTAACAT	TAACAGAACA	GTTAGCCAGA	TACGCCAGGG	GCATTACCTC	ATAAAGGACA	420
	AAGTGAGTTC	CACGCGTGCG	CCGTTTTAGA	TTAGTGATCC	CCCGGCTGCA	GGATTGATN	480
	GGGAGACAGT	CACGAGTCNN	GGACCACGTC	GGNGACCCAG	GCCCCAGNNT	GTGTCCNCCC	540
10	AGCCCCCAG	TCATGACGTT	TGTGAGCACG	ACGAGTCTGC	GGCCGGGCTG	GGGGCGCGTC	600
	TTCTGTCGCG	TGGGCCATCA	CTTCCTGAAT	GGCTGCGGTG	CGCTGATCGC	CCGAGCTGGC	660
	GAAGGGCGCC	ACAACCAGCG	CGCGCTCCGT	CTGCAGGCC	TTCCACGTGT	CGTGGAGTTC	720
	CTGAACGAAC	TCGGCCACCC	GCTCGGGGCC	CGTGCGCGCG	CGCGCGGCTT	GATAGCCGGC	780
	CGAGAGGCGC	CGCCAGCGCG	CCAGGAAC TG	ACTCATGTAA	CAGAACCCGG	GGACCTGGTC	840
15	CCCCGACATC	AACTTTGACG	CCCTGGCGTG	GATGCCCCGAC	ACGATGGCCA	GGAACCCGTG	900
	GATTTCCCGC	CGCACGACGG	CCAGCACGTT	ACCCTCGTGC	GAGACCTGGG	CCGCCAGCTC	960
	GTGCGATACC	CCGAGGTGCG	CCGTCGTCTC	GGTGACGACG	GACCGCAGCC	CCGCGAGGGA	1020
	CGCGACCAGC	GCGCGCTTGG	CGTCGTGATA	CATGCCGCAG	TACTGGCTCA	CCGCGTCGCC	1080
	CATGCCTCG	GGGCGCCAGG	GCCCCAGGCG	CTCGTGGGCG	TCTGCGACCA	CGGCGTACAG	1140
20	GCGGTGCCCC	TCGCTCTCGA	ACCGGCACTC	AAAGAAGGCG	GCGAGCGTGC	GCATGTGCAG	1200
	CCGCGAGCAG	ACGATCGCGT	CCTCCAGCTG	GCGGACCAGG	GGGTCGGCGC	GCTCGGCAAA	1260
	CTCTGCGATC	ACCCCCCGGG	CCGCCAGGGC	GTACATGCTG	ATCAGCAGCA	GGCTGCTGCC	1320
	CACCTCGGGA	GGCTGGGGGG	GAGGCAGCTG	GACCGCGGGC	CGCAGCTGCT	CGACGGCCCC	1380
	CCTGGCGATC	ACGTACAGCT	CGCGCAGCAG	CTGCTCGATG	TTGTGCGCCA	TCTGCATCGT	1440
25	GGGCCCCGAG	CCGGCCCCGG	TGGCCGGTTC	GAGGAGGGTG	ATCAGCGCGC	CCAATTTTGT	1500
	GCGGTGCCCC	TCGACGGTGG	GGAGATAGCC	CAGGCCGAAG	TCGCGCGCCC	AGGCCAGCAC	1560
	CCGCGAGGCA	AACTCGATGG	GGCGGGGCAG	GTAGGCAGCG	TTGCACGTGG	CCCTCAGCGC	1620
	GTCCCCGACC	ACCAGGGCCA	GCACGTAAGG	GACGAACCCC	GGGTCGGCGA	GGACGTTGGG	1680
	GTGGATGCCC	TCCAGGGCCG	GGAAGCGGAT	CTTGTTGGCC	GCGGCCAGGT	GAACCGAGGG	1740
30	GGCGTGCGTA	GGCGGCCCGA	CGGGGAGCAG	CGCGGACAGC	GGCGTGCCG	GGGTGGTGGG	1800
	GGTCAGGTCC	CAGTGGGTCT	GGCCGTACAC	GTCGAGCCAG	ATGAGCGCCG	TCTCGCGCAG	1860
	GAGGCTGGGC	TGGCCGGCGC	TGAAGCGGCG	CTCGGCCGTC	TCAAACCTCC	CCACGAGCGT	1920
	GCGCCGAGG	CTCGCCAGGT	GTTCCGTCGG	CACGCCCGGG	CCCATGATGC	GCGCCAGCGT	1980
	CTGGCTGAGG	ACGCCGCCCC	ACAGGCCGAC	CGCCTCACAG	AGCCGCCCCG	GCGTGTGCTC	2040
35	GCTGGCGCCC	TGGATCCGCC	GGAACGTTTT	CACGTAGCCG	GCGTAGTGCC	CGTACTCCCG	2100
	CGCGAGCCCC	AACACGTTTC	CCCCCGCAAG	GGCAATGCAC	CCAAAGAGCT	GCTGGATCTC	2160
	GCTGAGCCCC	TGGCCGGGGG	GCGTCCGCGC	GGGCACCCCC	GCCACCAAAA	ACCCCTCCAG	2220
	GGCCGATATG	TACTGGGTGC	AGTGCGCGGG	CGTGAACCCC	GCGTCGGTAA	GCGTGTGAT	2280
	CACCACGGAG	GGCGAGTTGC	TGTTTTGGAC	CAAAGCCAC	GTCTGCTGCA	GCAGCGCGAG	2340
40	GAGCCGTTGC	TGGGCCCCCG	CGGAGGGCGG	CTCCCTAGC	TGCAGCAGGC	CGGTGACGGC	2400
	CGGACGGAAG	ATGGCCAGCG	CCGACGCACT	CAGAAACGGC	ACGTGGGGGT	CGAAGACGGC	2460
	CGCGTCCGTC	CGCACGCGCG	CCATCAGCGT	CCCCGGGGGG	GCGCACGCGG	ACCGCGGGCT	2520
	GACGCGGCTT	AGGGCGGTTC	ACACGCGCAC	CTCCTCGCGA	CTGCGAACCA	TTTTGGTGGC	2580

CTCGAGGGGC GGGATCATGA TAGCCGGGTC GATCTCCCGC ACCGTGTGCT GAAACTGGGC 2640  
 CAGCAGCGGC GCGGGGACCA CCGCGCCCCG ATCGGGGGTC GTCAGGTACT CGTCCACCAG 2700  
 CGCCAGCGTA AACAGGGCCC GCGTGAGGGG GGTACAGGCG GCGTCGTCGA TGCCTGTAG 2760  
 GTGCGCCGAG AACAGCGTCA CCAATTGCT GACCAGGGCC AAGAACCGGA GACCTCTTG 2820  
 5 CACGATCGGG GACGGGAAGA GCAGGCTGTA CGCCGGGGTG GTCAGGTTGG CGCCGGGTTG 2880  
 CCCCAGGGGA ACCGGGGACA TCTTAAGCGA CATCTCCCCG AGGGCCTCCA GGGAGGTCCG 2940  
 CGGGTTTCATG GCCAGGCAGC TCTGGGTGAC GGTCCGCCAG CGGTCGATCC ACTCCACGGC 3000  
 AACTGGCGG ACGCGACCG GCCCAGGGC CGCGTGGTG CGCAGCCCG CGGCCTCCAG 3060  
 CGCGTGGGTC GTGTCGGAGC CGGTGATCGC CAGGACCGTG TCCTTGATGA CGTCCATCTC 3120  
 10 CCGGAAGGCC GCCTCGGGG TCTCGGGGAG CGCCACCGCC ATGCGGTGCA CCAGCAGCCC 3180  
 GGGGAGGTTT TCGGCCAAGA GCGCCGTCTC CGGAAGCCCG TGGGCCCGGT GCAAGGCGCA 3240  
 CAGTTGCTCC AGGAGCGGGT GCCAGCACGC CCGCGCCTCC GCCGGGCCGA CCGCCGCGCC 3300  
 CGACAACAGA AACGCCGCCG TGGCGGCGCG CAGTTTGGCC GCGGACAGAA ACGCCGGCTC 3360  
 GTCCGCGCTG CCCCGCGGCT CGCTCGAGGG GGAGGGCGGC CGGCGGAGGT TGGTCAGGCT 3420  
 15 CCCC AACAGG AECTGCAACG GTCCGTTTGG GGGTGGAGCG GACGGGGGGG TCATGCCGGC 3480  
 GGGCGCCGGG ACCTGGAGCG CGCTGTCCGA CATGGCGACC GGCCTGCGCG CTCGGCGACG 3540  
 CGGCGCGGAG ACCGCGGGCC CAAACGGGAA TGACTGCCGC CGCCCTATAC GGAGGGGGTA 3600  
 AGTATCGCCC GGGGACCCTT CGAAACCCCG GCGTGTCGC AAGTACGCC GCGAAAGGCG 3660  
 CGG 3664

20

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

25

- (A) LENGTH: 1043 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

Pro Arg Leu Ser Arg Ala Tyr Leu Arg His Arg Phe Glu Gly Ser Pro  
 1 5 10 15  
 35 Gly Asp Thr Tyr Pro Leu Arg Ile Gly Arg Arg Gln Ser Phe Pro Phe  
 20 25 30  
 Gly Pro Ala Val Ser Ala Pro Arg Arg Arg Ala Arg Thr Pro Val Ala  
 35 40 45  
 Met Ser Asp Ser Ala Leu Gln Val Pro Ala Pro Ala Gly Met Thr Pro  
 40 50 55 60  
 Pro Ser Ala Pro Pro Pro Asn Gly Pro Leu Gln Val Leu Leu Gly Ser  
 65 70 75 80  
 Leu Thr Asn Leu Arg Arg Pro Pro Ser Pro Ser Ser Glu Pro Ala Gly

		85		90		95										
	Ser	Ala	Asp	Glu	Pro	Ala	Phe	Leu	Ser	Ala	Ala	Lys	Leu	Arg	Ala	Ala
		100				105							110			
	Thr	Ala	Ala	Phe	Leu	Leu	Ser	Gly	Ala	Ala	Val	Gly	Pro	Ala	Glu	Ala
5		115				120							125			
	Arg	Ala	Cys	Trp	His	Pro	Leu	Leu	Glu	Gln	Leu	Cys	Ala	Leu	His	Arg
		130				135							140			
	Ala	His	Gly	Leu	Pro	Glu	Thr	Ala	Leu	Leu	Ala	Glu	Asn	Leu	Pro	Gly
	145				150						155				160	
10	Leu	Leu	Val	His	Arg	Met	Ala	Val	Pro	Glu	Thr	Pro	Glu	Ala	Ala	Phe
				165						170					175	
	Arg	Glu	Met	Asp	Val	Ile	Lys	Asp	Thr	Val	Leu	Ala	Ile	Thr	Gly	Ser
		180						185							190	
	Asp	Thr	Thr	His	Ala	Leu	Glu	Ala	Ala	Gly	Leu	Arg	Thr	Thr	Ala	Ala
15		195						200					205			
	Leu	Gly	Pro	Val	Arg	Val	Arg	Gln	Cys	Ala	Val	Glu	Trp	Ile	Asp	Arg
		210					215					220				
	Trp	Arg	Thr	Val	Thr	Gln	Ser	Cys	Leu	Ala	Met	Asn	Pro	Arg	Thr	Ser
	225				230						235				240	
20	Leu	Glu	Ala	Leu	Gly	Glu	Met	Ser	Leu	Lys	Met	Ser	Pro	Val	Pro	Leu
				245						250					255	
	Gly	Gln	Pro	Gly	Ala	Asn	Leu	Thr	Thr	Pro	Ala	Tyr	Ser	Leu	Leu	Phe
		260						265						270		
	Pro	Ser	Pro	Ile	Val	Gln	Glu	Gly	Leu	Arg	Phe	Leu	Ala	Leu	Val	Ser
25		275						280					285			
	Asn	Trp	Val	Thr	Leu	Phe	Ser	Ala	His	Leu	Gln	Arg	Ile	Asp	Asp	Ala
		290					295					300				
	Ala	Leu	Thr	Pro	Leu	Thr	Arg	Ala	Leu	Phe	Thr	Leu	Ala	Leu	Val	Asp
	305				310						315				320	
30	Glu	Tyr	Leu	Thr	Thr	Pro	Asp	Arg	Gly	Ala	Val	Val	Pro	Pro	Pro	Leu
				325						330					335	
	Leu	Ala	Gln	Phe	Gln	His	Thr	Val	Arg	Glu	Ile	Asp	Pro	Ala	Ile	Met
		340						345						350		
	Ile	Pro	Pro	Leu	Glu	Ala	Thr	Lys	Met	Val	Arg	Ser	Arg	Glu	Glu	Val
35		355						360					365			
	Arg	Val	Ser	Thr	Ala	Leu	Ser	Arg	Val	Ser	Pro	Arg	Ser	Ala	Cys	Ala
		370					375					380				
	Pro	Pro	Gly	Thr	Leu	Met	Ala	Arg	Val	Arg	Thr	Asp	Ala	Ala	Val	Phe
	385				390					395					400	
40	Asp	Pro	Asp	Val	Pro	Phe	Leu	Ser	Ala	Ser	Ala	Ile	Phe	Arg	Pro	Ala
				405						410					415	
	Val	Thr	Gly	Leu	Leu	Gln	Leu	Gly	Glu	Pro	Pro	Ser	Ala	Gly	Ala	Gln
		420						425					430			



Gln Arg Leu Leu Ala Leu Leu Gln Gln Thr Trp Ala Leu Val Gln Asn  
 435 440 445  
 Ser Asn Ser Pro Ser Val Val Ile Asn Thr Leu Thr Asp Ala Gly Phe  
 450 455 460  
 5 Thr Pro Ala His Cys Thr Gln Tyr Ile Ser Ala Leu Glu Gly Phe Leu  
 465 470 475 480  
 Val Ala Gly Val Pro Ala Arg Thr Pro Pro Gly His Gly Leu Ser Glu  
 485 490 495  
 10 Ile Gln Gln Leu Phe Gly Cys Ile Ala Gly Ala Asn Val Phe Gly Leu  
 500 505 510  
 Ala Arg Glu Tyr Gly His Tyr Ala Gly Tyr Val Lys Thr Phe Arg Arg  
 515 520 525  
 Ile Gln Gly Ala Ser Glu His Thr His Gly Arg Leu Cys Glu Ala Val  
 530 535 540  
 15 Gly Leu Ser Gly Gly Val Leu Ser Gln Thr Leu Ala Arg Ile Met Gly  
 545 550 555 560  
 Pro Ala Val Pro Thr Glu His Leu Ala Ser Leu Arg Arg Thr Leu Val  
 565 570 575  
 20 Gly Glu Phe Glu Thr Ala Glu Arg Arg Phe Ser Ala Gly Gln Pro Ser  
 580 585 590  
 Leu Leu Arg Glu Thr Ala Leu Ile Trp Leu Asp Val Tyr Gly Gln Thr  
 595 600 605  
 His Trp Asp Leu Thr Pro Thr Thr Pro Ala Thr Pro Leu Ser Ala Leu  
 610 615 620  
 25 Leu Pro Val Gly Pro Pro Ser His Ala Pro Ser Val His Leu Ala Ala  
 625 630 635 640  
 Ala Thr Lys Ile Arg Phe Pro Ala Leu Glu Gly Ile His Pro Asn Val  
 645 650 655  
 30 Leu Ala Asp Pro Gly Phe Val Pro Tyr Val Leu Ala Leu Val Val Gly  
 660 665 670  
 Asp Ala Leu Arg Ala Thr Cys Asn Ala Ala Tyr Leu Pro Arg Pro Ile  
 675 680 685  
 Glu Phe Ala Leu Arg Val Leu Ala Trp Ala Arg Asp Phe Gly Leu Gly  
 690 695 700  
 35 Tyr Leu Pro Thr Val Glu Gly His Arg Thr Lys Leu Gly Ala Leu Ile  
 705 710 715 720  
 Thr Leu Leu Glu Pro Ala Thr Arg Ala Gly Val Gly Pro Thr Met Gln  
 725 730 735  
 40 Met Ala Asp Asn Ile Glu Gln Leu Leu Arg Glu Leu Tyr Val Ile Arg  
 740 745 750  
 Ala Val Glu Gln Leu Arg Pro Ala Val Gln Leu Pro Pro Pro Gln Pro  
 755 760 765  
 Pro Glu Val Gly Ser Ser Leu Leu Leu Ile Ser Met Tyr Ala Arg Val

	770		775		780											
	Met	Gln	Glu	Phe	Ala	Glu	Arg	Ala	Asp	Pro	Leu	Val	Arg	Gln	Leu	Glu
	785				790					795					800	
	Asp	Ala	Ile	Val	Leu	Leu	Arg	Leu	His	Met	Arg	Thr	Leu	Ala	Ala	Phe
5				805						810					815	
	Phe	Glu	Cys	Arg	Phe	Glu	Ser	Asp	Gly	His	Arg	Leu	Tyr	Ala	Val	Val
				820						825					830	
	Ala	Asp	Ala	His	Glu	Arg	Leu	Gly	Pro	Trp	Arg	Pro	Glu	Ala	Met	Gly
				835					840					845		
10	Asp	Ala	Val	Ser	Gln	Tyr	Cys	Gly	Met	Tyr	His	Asp	Ala	Lys	Arg	Ala
				850					855					860		
	Leu	Val	Ala	Ser	Leu	Ala	Gly	Leu	Arg	Ser	Val	Val	Thr	Glu	Thr	Thr
				865					870					875		880
	Ala	His	Leu	Gly	Val	Cys	Asp	Glu	Leu	Ala	Ala	Gln	Val	Ser	His	Glu
15				885						890					895	
	Gly	Asn	Val	Leu	Ala	Val	Val	Arg	Arg	Glu	Ile	His	Gly	Phe	Leu	Ala
				900						905					910	
	Ile	Val	Ser	Gly	Ile	His	Ala	Arg	Ala	Ser	Lys	Leu	Met	Ser	Gly	Asp
				915						920					925	
20	Gln	Val	Pro	Gly	Phe	Cys	Tyr	Met	Ser	Gln	Phe	Leu	Ala	Arg	Trp	Arg
				930						935					940	
	Arg	Leu	Ser	Ala	Gly	Tyr	Gln	Ala	Ala	Arg	Ala	Ala	Thr	Gly	Pro	Glu
				945						950					955	960
	Arg	Val	Ala	Glu	Phe	Val	Gln	Glu	Leu	His	Asp	Thr	Trp	Lys	Gly	Leu
25				965						970					975	
	Gln	Thr	Glu	Arg	Ala	Leu	Val	Val	Ala	Pro	Phe	Ala	Ser	Ser	Gly	Asp
				980						985					990	
	Gln	Arg	Thr	Ala	Ala	Ile	Gln	Glu	Val	Met	Ala	His	Ala	Asn	Glu	Asp
				995						1000					1005	
30	Ala	Pro	Pro	Ala	Arg	Pro	Gln	Thr	Arg	Arg	Ala	His	Lys	Arg	His	Asp
				1010						1015					1020	
	Trp	Gly	Ala	Gly	Xaa	Thr	Xaa	Xaa	Gly	Ala	Trp	Val	Xaa	Asp	Val	Val
				1025						1030					1035	104
	Xaa	Asp	Ser													

## (2) INFORMATION FOR SEQ ID NO:87:

## (i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 5033 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

5	CGCGGTGCGAC TCTAGAGGAT CCCCTGCGCC GCGTCGGGAT TCACCAACTC GTTCGCGCGC	60
	TGCAGGAGGT TCTTGCCCTC GCAGACCGTC ACGCGAATGG TGGTGAGGTC GAGGAGCTCG	120
	TTGAGGTCTC CGTCGGTGTG CGGCCGCGAC ATGTCCCACA GCTGTACCGC CGCCAGCCGG	180
	GCGTGCGTGG CCGCCAGGCG CCCGACCGCG GCGCAAAAAA CGCGCTTGTT GAACCCGGCC	240
	ACCCGGGGGG TCCACGGCGC CGTGGGGCTC GGTGGGGCGG TGCTGAATTG CACCTCCTTG	300
10	GCCAGTCCCT GGGCGGGTGT CTTGGTTCTT CCCGAGGCCG TGGGAGCGGG GCGGTCTAGG	360
	AGCACGGCGG TATCGGCCTG GCGGGGTCGC CTGCCGCGGG CAGGGTCGGT CGCCGGGGTC	420
	GCGGGGGCCT TAGGGCGCCC CGCGCGTCAT TTTGGGGGTC CGCGCGGGAG GGGCGTGC GA	480
	GCGCCCCCGG GCGCCCCACG GCGCCCCGGG GGGTGGAGGA GCGCGCGCGG GCGCGGGGCC	540
	GTGAGAGCCC GCGACGGACG CCGAACGACG CGGTCGCGCG GTATCCCGGG ACTCGTCGTT	600
15	GTCTTCGGAC GACGACGAGT CCCGGTAGAG GGCATACCCA GCCTCGTCAT AATGGAGAAA	660
	GCGAACCTCG CCCCTTGGGC GCGCGCGCAT CCGGCCAGCG CCGCGCGGGA AGTCGTCGCG	720
	CGGACTCTCT GGATCCGCCG GGGAGACCGG GCCATAGTAC AGCTCCTCGT GGGTCCCGCG	780
	CGGCGCTTCC CGCGGACACG ACTTGACGGA GCGGCGAGAG GTCATGGTCT ATCGGAGACA	840
	CCGGGGACGC CCGTGC GGAT CACAGGGAAG GCGTCGGCGA AGCAGGCAGA GAGCGTCGGA	900
20	AGGCGGCGAG GGAGGGAAAG AGGGAGACCG GCGGGGTACG GGAGAGCAGC GAGGGCCTGC	960
	GTAACCCACG GGGGCCGCGG GAGTGGCTCC CTGCGGGTTC GGGGGGAGAG TTTATAGGAA	1020
	GTGGATATAA CCGCAGGCGA CCGGACTAAC CAATCCCCCG GGGGGCAACG GACAGACACG	1080
	CCCCGAACCG GCGCGACTTC CCGGAGGAAG CAAAGGCCCG GGGCCGCCCA ACGACACGCC	1140
	CACCCCTTCC CAACAGGGCG GGCTCAGGCT GACCCGGCGG CCAGTGCCCG CTGGCATATC	1200
25	TGATACACGT GCGCGATCAT ACATACGCCC ATCGAGGTCA TGCCCTAGATA AAAGGGCACC	1260
	AGGACCCCCG GGACGGACAC CACACCGGCG CTGTGCGCCC GGCATTGCGC GTCCCCGATA	1320
	ACGCCGCGTG CGCCTGCCGC GTTCGGCGGC TCCCCGGGCA CGCCCGCGAC GAGCGCGACG	1380
	AAACACAGCA CCACCCAGCG GCCCAGTCTT GCGGGTTTCC CCGTCATCGC GGCGATGAGT	1440
	CAGTGGGGGC CCAGGGCGAT CCTTGTCCAG ACGGACAGCA CCAACCGGAA TGCCGATGGG	1500
30	GACTGGCAAG CGGCCGTAGC TATTCGCGGG GCGGAGTCG TTCAACTGAA CATGGTCAAC	1560
	AAACGCGCCG TGGATTTTAC CCCGGCAGAA TGCGGGGACT CCGAATGGGC CGTGGGCCGC	1620
	GTCTCTCTGG GCCTGCGAAT GGCAATGCCG CGTGACTTCT GCGCGATTAT TCACGCCCCC	1680
	GCGGTATCCG GCGCCGGGCC CCACGTGATG CTCGGTCTCG TCCACTCGGG CTACCGCGGA	1740
	ACCGTCCTGG CCGTGGTCGT ATCCCCGAAC GGGACGCGCG GGTTCGCCCG CGGGGCCCTC	1800
35	CGGGTCGACG TGACGTTTCT GGACATCCGG GCCACCCCCC CGACCCTCAC CGAGCCGAGC	1860
	TCCCTGCACC GGTTCGCGA GTTGGCGCCG TCCCCGCTGG CAGGGTTACG AGAAGATCCT	1920
	TGGTTGGACG GGGCGCTCGC GACCGCCGGG GGGGCGGTGG CCCTGCCGGC CAGACGGCGC	1980
	GGGGGATCGC TGGTCTACGC GGGCGAGCTA ACGCAGGTGA CCACCGAGCA CGGCGACTGC	2040
	GTGCACGAGG CGCCCGCCTT TCTGCCAAAG CGCGAGGAGG ACGCAGGCTT TGACATTCTC	2100
40	ATCCACCGAG CCGTGACCGT CCCGGCCAAC GCGGCCACGG TCATACAGCC GTCCCTCCGC	2160
	GTATTGCGCG CGGCCGACGG ACCAGAGGCC TGCTATGTGC TGGGGCGGTC GTCGCTCAAT	2220
	GCCAGGGGCC TCCTGGTCAT GCCTACGCGC TGGCCCTCCG GGCACGCCTG TGCGTTTGTT	2280
	GTATGTAACC TGACCGGAGT CCCGGTGACC CTACAAGCCG GGTCCAAGGT CGCCCAGCTG	2340

	CTCGTCGCGG	GGACCCACGC	CCTCCCCTGG	ATCCCCCCCCG	ACAACATCCA	CGAGGACGGC	2400
	GCATTCCGGG	CCTACCCCAG	AGGGGTTCG	GACGCGACCG	CCACCCCCCG	AGACCCGCGG	2460
	ATTTTGGTGT	TTACGAACGA	GTTTGACGCG	GACGCCCCCG	CAAGCAAGCG	GGGGGCCGGG	2520
	GGGTTTGGCT	CCACTGGCAT	CTAAACCGCG	CCTCGCGTCG	GGCCAGATGG	GGCCCCGGTC	2580
5	AATAAAGAGC	TCTGTTTCGC	ATATGCCCTG	GTGTTGGCGG	TTTTTTTTTTT	GTTGTCTGTC	2640
	TGCCCCGGAC	TCGGTTGTCC	GTTCTGTCTG	CGCTATCACA	TACGCACAAA	CACACGGGTA	2700
	GAGTGGAACC	GAAACCGGTC	GACGTTTATT	CACCACACAG	AAACACAAGC	TAAGCGAGAA	2760
	GGAGGGGGGC	CTCGGTCGAC	GAGGCCTGGC	GTTTGGGGGC	GGACGTGCGA	TGACGTGGGT	2820
	CCGGTGTAGG	GTCCGCGGGG	GGCACGGGCC	CGGGGCGAAC	GGGGGATCTG	TCGCCGCGCT	2880
10	GGGTGACTGG	GACCGACGCA	ACCTCCGGGG	CTTGTGCCCT	CGTAGGCCCCG	GGGGGGGCCCT	2940
	CGGTGCGTCC	GAGCCCCGCG	GTGCGGGTCC	CTCCGGCCAG	AGCCGAGGTG	GAGAGACCAA	3000
	GGGCCCCGTC	CGCGATCGCC	ACGTCTCTCA	TGACCACGTC	GCTTTTCGGCC	ATGCTCCGAA	3060
	TGGCCTGGGA	GACGAGCACG	TCCGCCGACT	TGTCCGCGGC	CCCCACCGAC	ATGTACATCT	3120
	GCAGGATGGT	GGCCATGCAC	GTGTCCGCCA	GGCGGCGCAT	CTTGTCCCGA	TGCGCCGCAA	3180
15	CGGCCCCGTC	GATGGTGGAG	CCCTCGAGTC	CCGGGTGGTG	GCGCGCCAGC	CTCTCGAGGT	3240
	TGACCATGCA	GGCGTGGTAT	GTGCGGGCCA	GGGCGCGCGC	CTTCACGAGG	CGCCGGGTGT	3300
	CGTCCAGCGA	CTCTAGGGCG	TCATCAAGCG	TGATGGGGGC	GGGCAAAAGC	GCATTGACCA	3360
	CCGCCAGGGC	CTCCTGCAGC	CGCGGCTCCG	CCTCCGAGGG	CGGATCCGCG	GCCCCGAATCA	3420
	TCTCATATTG	TTGTTCTCTG	GGGCGCGTGC	CCCAACCGCA	CAGCACCCCG	AGCAGGGACG	3480
20	CCATCCCGBA	ACACGCGCGC	GGCTCTGCGC	CGGCTTTCCC	CCACCCACC	CCCTCCGGGT	3540
	TCGAGGGGCG	GATGGGGACG	GAAGACTGCG	ATCACGAAGG	GCGGTGCGTT	GCGGCTCCCG	3600
	TGGAGGTTAT	GGCGCTGTAT	GCGACCGACG	GGTGCCTTAT	CACCTCCTCG	CTCGCCCTCC	3660
	TCACAAACTG	CCTGCTGGGG	GCCGAGCCGT	TGTATATATT	CAGCTACGAC	GCGTACCGGC	3720
	CCGATGCGCC	CAATGGCCCC	ACGGGCGCGC	CCACCGAACA	GGAGAGGTTT	GAGGGGAGCC	3780
25	GGGCGCTCTA	CCGGGATGCG	GGGGGGCTAA	ATGGCGATTG	ATTTTCGGGTG	ACCTTTTGTG	3840
	TATTGGGGAC	GGAAGTGGGC	GTGACCCACC	ACCCGAAAGG	GCGCACCCCG	CCCATGTTTG	3900
	TGTGCCGCTT	CGAGCGAGCG	GACGACGTCG	CCGTGCTCCA	AGACGCCCTG	GGCCGCGGGA	3960
	CCCCATTGCT	CCCGGCCAC	ATCACAGCAA	CTCTGGACTT	GGAGGCGACG	TTTGCGCTCC	4020
	ACGCTAACAT	CATCATGGCT	CTCACCGTGG	CCATAGTCCA	CAACGCCCCC	GCCCGCATCG	4080
30	GCAGCGGCAG	CACCGCTCCC	CTGTATGAGC	CCGGCGAATC	GATGCGCTCG	GTCGTGCGGC	4140
	GCATGTCCCT	GGGGCAGCGC	GGCCTCACCA	CGCTGTTCTG	GCACCACAAG	GCGCGCGTGC	4200
	TGGCGGCGTA	CCGCCGGGCG	TATTATGGGA	GCGCCCAAAG	CCCCTTTGTG	TTTCTGAGCA	4260
	AATTTCGGCC	GGACAAAAAG	AGCCTGGTGC	TGGCCGCTAG	GTACTACCTA	CTCCAGGCTC	4320
	CGCGCTTGGG	GGGCGCCGGA	GCCACGTACG	ATCTGCAGGC	CGTGAAAGAC	ATCTGCGCGA	4380
35	CCTACGCGAT	CCCCACGAC	CCACGCCCCG	ACACCTCAG	TGCCGCGTCC	TTGACCTCGT	4440
	TCGCCGCCAT	CACTCGGTTC	TGTTGCACGA	GCCAGTACTC	CCGCGGGGCC	GCGGCCGCTG	4500
	GGTTTCCGCT	GTATGTGGAG	CGCCGCATCG	CCGCCGACGT	ACGCGAGACC	GGCGCGCTGG	4560
	AGAAGTTCAT	CGCCCACGAT	CGCAGTTGCC	TGCGCGTGTC	CGACCGGGAA	TTCATTACGT	4620
	ACATCTACCT	GGCCCACTTT	GAGTGCTTCA	GCCCCCGCG	CCTGGCCACG	CATCTCCGGG	4680
40	CCGTGACCAC	CCACGACCCC	AGCCCCGCGG	CCAGCACGGA	GCAGCCCTCG	CCCCTGGGTC	4740
	GGGAGGCGGT	GGAACAGTTC	TTCCGGCACG	TGCGCGCCCA	GCTGAACATC	CGCGAGTACG	4800
	TAAAGCAAAA	CGTCACCCCC	AGGGAAACCG	CCCTGGCGGG	AGACGCGGCC	GCCGCCTACC	4860
	TGCGCGCGCG	CACGTATGCC	CCGGCGGCCC	TCACGCCCGC	CCCCGCGTAC	TGCGGGGTGC	4920

CAGACTCGTC CACCAAAATG ATGGGACGTC TGGCGGAAGC AGAAAGGCTC CTAGTCCCCC 4980  
ACGGCTGGCC CGCGTTCGCA CCAACAACCC CCGGGGACGA CGCGGGGGGC GG 5033

## (2) INFORMATION FOR SEQ ID NO:88:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 117 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

15

Val Leu Leu Asp Ala Pro Ala Pro Thr Ala Ser Gly Arg Thr Lys Thr  
1 5 10 15  
Pro Ala Gln Gly Leu Ala Lys Glu Val Gln Phe Ser Thr Ala Pro Pro  
20 25 30  
Ser Pro Thr Ala Pro Trp Thr Pro Arg Val Ala Gly Phe Asn Lys Arg  
35 40 45  
Val Phe Cys Ala Ala Val Gly Arg Leu Ala Ala Thr His Ala Arg Leu  
50 55 60  
Ala Ala Val Gln Leu Trp Asp Met Ser Arg Pro His Thr Asp Gly Asp  
25 65 70 75 80  
Leu Asn Glu Leu Leu Asp Leu Thr Thr Ile Arg Val Thr Val Cys Glu  
85 90 95  
Gly Lys Asn Leu Leu Gln Arg Ala Asn Glu Leu Val Asn Pro Asp Ala  
100 105 110  
30 Ala Gln Gly Ile Leu  
115

## (2) INFORMATION FOR SEQ ID NO:89:

35

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 131 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

40

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Met Thr Ser Arg Arg Ser Val Lys Ser Cys Pro Arg Glu Ala Pro Arg  
 1 5 10 15  
 Gly Thr His Glu Glu Leu Tyr Tyr Gly Pro Val Ser Pro Ala Asp Pro  
 5 20 25 30  
 Glu Ser Pro Arg Asp Asp Phe Arg Arg Gly Ala Gly Pro Met Arg Ala  
 35 40 45  
 Arg Pro Arg Gly Glu Val Arg Phe Leu His Tyr Asp Glu Ala Gly Tyr  
 50 55 60  
 10 Ala Leu Tyr Arg Asp Ser Ser Ser Ser Glu Asp Asn Asp Glu Ser Arg  
 65 70 75 80  
 Asp Thr Ala Arg Pro Arg Arg Ser Ala Ser Val Ala Gly Ser His Gly  
 85 90 95  
 Pro Gly Pro Ala Arg Ala Pro Pro Pro Gly Gly Pro Val Gly Ala  
 15 100 105 110  
 Gly Gly Arg Ser His Ala Pro Pro Ala Arg Thr Pro Lys Met Thr Arg  
 115 120 125  
 Gly Ala Pro  
 130  
 20

## (2) INFORMATION FOR SEQ ID NO:90:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 363 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

Met Ser Gln Trp Gly Pro Arg Ala Ile Leu Val Gln Thr Asp Ser Thr  
 1 5 10 15  
 35 Asn Arg Asn Ala Asp Gly Asp Trp Gln Ala Ala Val Ala Ile Arg Gly  
 20 25 30  
 Gly Gly Val Val Gln Leu Asn Met Val Asn Lys Arg Ala Val Asp Phe  
 35 40 45  
 Thr Pro Ala Glu Cys Gly Asp Ser Glu Trp Ala Val Gly Arg Val Ser  
 40 50 55 60  
 Leu Gly Leu Arg Met Ala Met Pro Arg Asp Phe Cys Ala Ile Ile His  
 65 70 75 80  
 Ala Pro Ala Val Ser Gly Pro Gly Pro His Val Met Leu Gly Leu Val

	85	90	95
	His Ser Gly Tyr Arg Gly Thr Val Leu Ala Val Val Val Ser Pro Asn		
	100	105	110
	Gly Thr Arg Gly Phe Ala Pro Gly Ala Leu Arg Val Asp Val Thr Phe		
5	115	120	125
	Leu Asp Ile Arg Ala Thr Pro Pro Thr Leu Thr Glu Pro Ser Ser Leu		
	130	135	140
	His Arg Phe Pro Gln Leu Ala Pro Ser Pro Leu Ala Gly Leu Arg Glu		
	145	150	155
10	160		
	Asp Pro Trp Leu Asp Gly Ala Thr Ala Gly Gly Ala Val Pro Ala Arg		
	165	170	175
	Arg Arg Gly Gly Ser Leu Val Tyr Ala Gly Glu Leu Thr Gln Val Thr		
	180	185	190
	Thr Glu His Gly Asp Cys Val His Glu Ala Pro Ala Phe Leu Pro Lys		
15	195	200	205
	Arg Glu Glu Asp Ala Gly Phe Asp Ile Leu Ile His Arg Ala Val Thr		
	210	215	220
	Val Pro Ala Asn Gly Ala Thr Val Ile Gln Pro Ser Leu Arg Val Leu		
	225	230	235
20	240		
	Arg Ala Ala Asp Gly Pro Glu Ala Cys Tyr Val Leu Gly Arg Ser Ser		
	245	250	255
	Leu Asn Arg Leu Leu Val Met Pro Thr Arg Trp Pro Ser Gly His Ala		
	260	265	270
	Cys Ala Phe Val Val Cys Asn Leu Thr Gly Val Pro Val Thr Leu Gln		
25	275	280	285
	Ala Gly Ser Lys Val Ala Gln Leu Leu Val Ala Gly Thr His Ala Leu		
	290	295	300
	Pro Trp Ile Pro Pro Asp Asn Ile His Glu Asp Gly Ala Phe Arg Ala		
	305	310	315
30	320		
	Tyr Pro Arg Gly Val Pro Asp Ala Thr Ala Thr Pro Arg Asp Pro Pro		
	325	330	335
	Ile Leu Val Phe Thr Asn Glu Phe Asp Ala Asp Ala Pro Pro Ser Lys		
	340	345	350
	Arg Gly Ala Gly Gly Phe Gly Ser Thr Gly Ile		
35	355	360	

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:

40

(A) LENGTH: 251 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

5  
Val Gly Trp Gly Lys Ala Gly Ala Glu Pro Arg Ala Cys Ser Gly Met  
1 5 10 15  
Ala Ser Leu Leu Gly Val Leu Cys Gly Trp Gly Trp Glu Glu Gln Gln  
20 25 30  
10 Tyr Glu Met Ile Arg Ala Ala Asp Pro Pro Ser Glu Ala Glu Pro Arg  
35 40 45  
Leu Gln Glu Ala Val Val Asn Ala Leu Leu Pro Ala Pro Ile Thr Leu  
50 55 60  
Asp Asp Ala Leu Glu Ser Leu Asp Asp Thr Arg Arg Leu Val Lys Ala  
15 65 70 75 80  
Arg Ala Arg Thr Tyr His Ala Cys Met Val Asn Leu Glu Arg Leu Ala  
85 90 95  
Arg His His Pro Gly Leu Glu Gly Ser Thr Ile Asp Gly Ala Val Ala  
100 105 110  
20 Ala His Arg Asp Lys Met Arg Arg Leu Ala Asp Thr Cys Met Ala Thr  
115 120 125  
Ile Leu Gln Met Tyr Met Ser Val Gly Ala Ala Asp Lys Ser Ala Asp  
130 135 140  
Val Leu Val Ser Gln Ala Ile Arg Ser Met Ala Glu Ser Asp Val Val  
25 145 150 155 160  
Met Glu Asp Val Ala Ile Ala Glu Arg Ala Leu Gly Leu Ser Thr Ser  
165 170 175  
Ala Gly Gly Thr Arg Thr Ala Gly Leu Gly Ala Thr Glu Ala Pro Pro  
180 185 190  
30 Gly Pro Thr Arg Ala Gln Ala Pro Glu Val Ala Ser Val Pro Val Thr  
195 200 205  
His Ala Gly Asp Arg Ser Pro Val Arg Pro Gly Pro Val Pro Pro Ala  
210 215 220  
Asp Pro Thr Pro Asp Pro Arg His Arg Thr Ser Ala Pro Lys Arg Gln  
35 225 230 235 240  
Ala Ser Ser Thr Glu Ala Pro Leu Leu Leu Ala  
245 250

(2) INFORMATION FOR SEQ ID NO:92:

40

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 710 amino acids

(B) TYPE: amino acid



(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

	Val	Thr	Gly	Thr	Asp	Ala	Thr	Ser	Gly	Ala	Cys	Ala	Leu	Val	Gly	Pro
	1				5					10					15	
10	Gly	Gly	Ala	Ser	Val	Ala	Pro	Ser	Pro	Ala	Val	Arg	Val	Pro	Pro	Ala
				20					25					30		
	Arg	Ala	Glu	Val	Glu	Arg	Pro	Arg	Ala	Arg	Ser	Ala	Ile	Ala	Thr	Ser
		35					40					45				
	Ser	Met	Thr	Thr	Ser	Leu	Ser	Ala	Met	Leu	Arg	Met	Ala	Trp	Glu	Thr
15		50					55					60				
	Ser	Thr	Ser	Ala	Asp	Leu	Ser	Ala	Ala	Pro	Thr	Asp	Met	Tyr	Ile	Cys
	65				70					75					80	
	Arg	Met	Val	Ala	Met	His	Val	Ser	Ala	Arg	Arg	Arg	Ile	Leu	Ser	Arg
				85						90				95		
20	Cys	Ala	Ala	Thr	Ala	Pro	Ser	Met	Val	Glu	Pro	Ser	Ser	Pro	Gly	Trp
				100					105					110		
	Trp	Arg	Ala	Ser	Leu	Ser	Arg	Leu	Thr	Met	Gln	Ala	Trp	Tyr	Val	Arg
		115					120						125			
	Ala	Arg	Ala	Arg	Ala	Phe	Thr	Arg	Arg	Arg	Val	Ser	Ser	Ser	Asp	Ser
25		130				135					140					
	Arg	Ala	Ser	Ser	Ser	Val	Met	Gly	Ala	Gly	Lys	Ser	Ala	Leu	Thr	Thr
	145				150					155				160		
	Ala	Arg	Ala	Ser	Cys	Ser	Arg	Gly	Ser	Ala	Ser	Glu	Gly	Gly	Ser	Ala
				165					170					175		
30	Ala	Arg	Ile	Ile	Ser	Tyr	Cys	Cys	Ser	Ser	Gly	Arg	Val	Pro	Gln	Pro
			180				185						190			
	His	Ser	Thr	Pro	Ser	Arg	Asp	Ala	Ile	Pro	Glu	His	Arg	Ser	Ala	Pro
		195				200						205				
	Ala	Phe	Pro	His	Pro	Thr	Pro	Ser	Gly	Phe	Ala	Gly	Ala	Met	Gly	Thr
35		210				215						220				
	Glu	Asp	Cys	Asp	His	Glu	Gly	Arg	Ser	Val	Ala	Ala	Pro	Val	Glu	Val
	225				230					235				240		
	Met	Ala	Leu	Tyr	Ala	Thr	Asp	Gly	Cys	Val	Ile	Thr	Ser	Ser	Leu	Ala
				245					250					255		
40	Leu	Leu	Thr	Asn	Cys	Leu	Leu	Gly	Ala	Glu	Pro	Leu	Tyr	Ile	Phe	Ser
				260				265					270			
	Tyr	Asp	Ala	Tyr	Arg	Pro	Asp	Ala	Pro	Asn	Gly	Pro	Thr	Gly	Ala	Pro
		275				280						285				

Thr Glu Gln Glu Arg Phe Glu Gly Ser Arg Ala Leu Tyr Arg Asp Ala  
 290 295 300  
 Gly Gln Gly Asp Ser Phe Arg Val Thr Phe Cys Leu Leu Gly Thr Glu  
 305 310 315 320  
 5 Val Gly Val Thr His His Pro Lys Gly Arg Trp Met Phe Val Cys Arg  
 325 330 335  
 Phe Glu Arg Ala Asp Asp Val Ala Val Leu Gln Asp Ala Leu Gly Arg  
 340 345 350  
 Gly Thr Pro Leu Leu Pro Ala His Ile Thr Ala Thr Leu Asp Leu Glu  
 10 355 360 365  
 Ala Thr Phe Ala Leu His Ala Asn Ile Ile Met Ala Leu Thr Val Ala  
 370 375 380  
 Ile Val His Asn Ala Pro Ala Arg Ile Gly Ser Gly Ser Thr Ala Pro  
 385 390 395 400  
 15 Leu Tyr Glu Pro Gly Glu Ser Met Arg Ser Val Val Gly Arg Met Ser  
 405 410 415  
 Leu Gly Gln Arg Gly Leu Thr Thr Leu Phe Val His His Lys Ala Arg  
 420 425 430  
 Val Leu Ala Ala Tyr Arg Arg Ala Tyr Tyr Gly Ser Ala Gln Ser Pro  
 20 435 440 445  
 Phe Trp Phe Leu Ser Lys Phe Gly Pro Asp Lys Lys Ser Leu Val Leu  
 450 455 460  
 Ala Ala Arg Tyr Tyr Leu Leu Gln Ala Pro Arg Leu Gly Gly Ala Gly  
 465 470 475 480  
 25 Ala Thr Tyr Asp Leu Gln Ala Val Lys Asp Ile Cys Ala Thr Tyr Ala  
 485 490 495  
 Ile Pro His Asp Pro Arg Pro Asp Thr Leu Ser Ala Ala Ser Leu Thr  
 500 505 510  
 Ser Phe Ala Ala Ile Thr Arg Phe Cys Cys Thr Ser Gln Tyr Ser Arg  
 30 515 520 525  
 Gly Ala Ala Ala Ala Gly Phe Pro Leu Tyr Val Glu Arg Arg Ile Ala  
 530 535 540  
 Ala Asp Val Arg Glu Thr Gly Ala Leu Glu Lys Phe Ile Ala His Asp  
 545 550 555 560  
 35 Arg Ser Cys Leu Arg Val Ser Asp Arg Glu Phe Ile Thr Tyr Ile Tyr  
 565 570 575  
 Leu Ala His Phe Glu Cys Phe Ser Pro Pro Arg Leu Ala Thr His Leu  
 580 585 590  
 Arg Ala Val Thr Thr His Asp Pro Ser Pro Ala Ala Ser Thr Glu Gln  
 40 595 600 605  
 Pro Ser Pro Leu Gly Arg Glu Ala Val Glu Gln Phe Phe Arg His Val  
 610 615 620  
 Arg Ala Gln Leu Asn Ile Arg Glu Tyr Val Lys Gln Asn Val Thr Pro

625                      630                      635                      640  
 Arg Glu Thr Ala Gly Asp Ala Ala Ala Tyr Leu Arg Ala Arg Thr  
                          645                      650                      655  
 Tyr Ala Pro Ala Ala Leu Thr Pro Ala Pro Ala Tyr Cys Gly Val Ala  
 5                      660                      665                      670  
 Asp Ser Ser Thr Lys Met Met Gly Arg Leu Ala Glu Ala Glu Arg Leu  
                          675                      680                      685  
 Leu Val Pro His Gly Trp Pro Ala Phe Ala Pro Thr Thr Pro Gly Asp  
                          690                      695                      700  
 10 Asp Ala Gly Gly Gly Ile  
                          705                      710

## (2) INFORMATION FOR SEQ ID NO:93:

## 15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5742 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

25 AGAGGAGTAG AATAGAGAAG AGGATAGAGA GGAGTAGAGT GATCAATAAG ATGTGAAATA 60  
 TGAGAGAGTA GTAAAGTAGT AAAGAATTTC GGACGGAGCG TAGACAGATA GATATAGAGA 120  
 TATGCCCGTC TAGGAAGAAG AAATGTGTGA GAGATATAAA GTGGGTAAGA GGTCTATATG 180  
 AAGAGTAACG AGTAAGGGAT GGGTAGAAGA AGCCTGATGG GGAAGGTGAG AAGAAGTGTT 240  
 AAAGGGGATA GAATGGAGGT TAGCGAGGTG GTAGAAACAA GAAGGGGAAT AGGAAGGAAC 300  
 GGCCAATAGG AGAAAGAAGA GGAATGATGG AGGGAAGATG AGGTAGGAGC CATCCCGGCC 360  
 30 CACATTTACG GAAAACAGAC CAACGTGCAG GTCGCGACGG AGTTCGATAT GGAAGTAGAA 420  
 GTTCTCCGCG GCGCGGTCCC AAATCGGCAC CAGCAGGGAA GCATTTACAA AAGCGTACCG 480  
 GGTCCGAGGC CCGCCGCCCT GGTACGTGTA CGTGTAACA CCGCACGTCT TTGGGGACGG 540  
 CGGCTCGCCG CCGCGACACC CTCCATACTG CCGAAGGAGG ATGGGCTGAC GGCAGGCGCG 600  
 GGTGAGCCGG TATTACGTCA CACGGGCCGC ATACGTTGCG TTGGGTGCGC CAAATCCGCG 660  
 35 TACCGGCGGC GCCAGCAAAA CCAGCCCCC CGTGAGGAAC GAGCGGCCCA GGGGCTCGTG 720  
 ACGGACGACG CGCCGGGGCA GGTCTTGCCG CCCGGCGTCC GCGGAACCCA AGGGCCCGGG 780  
 ATCGTCCAAC CGGGGATAGG CATAACAATA TTAGCCACG GGGGAACCTC CCGGAAAAC 840  
 AACATCGTTG TTGGGGGGAT TAATTGGTCC GGGGACACCC GACCCGCCGC GTGTCCCCGG 900  
 AAGACCAGAA AGAACA AAAA GAAGAAGCAA CCTGGGAGCG ATGGCGTGCA TGACGCCGGG 960  
 40 CGGCAAGGTG CAAAAACAC CAAAGCCCGA GGGCCGCGTC TTTTGTGCA AAAACATCCA 1020  
 ACCAGCCCCC CCTCCACGCC CCCCAGGGGG AGCGGGGTCA CTTAGGGTGA AATAGCGGCA 1080  
 GGCGCAGCAA CTCCGCGGCG CTTGGGCGGA GCGCCGCGTC AAAGGTAAGG GCTTTGCAGA 1140  
 TGAGATATTC GACGTCTGTG TGGATCTTGT AGTAGCGGGT CCATGCCGGT CGGGTCCACG 1200

	CCGGACGATT	GTTCCCGGCC	CCCCGCGAGC	GGTAGTGCGC	GGTGAGGCGC	GATTCCGCGT	1260
	GCGTTGGAAA	CTCGTCGACG	TGTACCTGGG	CCTGTGGGAT	GATGCGCGCG	ATCTGGTTGT	1320
	CGCACGGCCG	CCTTTCGGGG	TCGCGCGGGG	CCGAGAACAA	GGACGCGGTG	TGGACGGCCG	1380
	TCTCAAAGAT	CACCAGGCCG	GCGCTCCAGA	TGTCGATTAC	CTGGGTGTAC	GGATCCCCGG	1440
5	CCAGGACCTC	GGGGGCGTTT	GTATCGATGG	TGCCTGCGAT	CCCCTAATGG	AAGGGGCTCG	1500
	ATCGACACCC	GCGCACAAAG	CACGCCGCCC	CAAAGTCCCC	CAGACAGATG	TTCTCGGGGG	1560
	TGTTGATGAG	GATGTTCTCG	GTCTTAATAT	CGCGGTGGAT	GATGCCTTCG	CAGTGGACGT	1620
	AGTCGATGGC	GCTCAAGAGC	TGCCGGGAGA	CCGCGGTTAT	CTGTAGGTGG	CCCAACGGAG	1680
	ACGGGCGCTT	GCTCAGATAG	GTATACAGGT	CGCAGTGATA	CTTGGGGAGG	ACCAGACACG	1740
10	TGACCCCAGA	AACGACGTGC	AGGTCCAGGA	GGGGTAGGAT	CGCGGGGTGG	TTCAGGCGTC	1800
	TCAGCAGCCG	CGCCTCGTGG	TTTGTGCTGG	CGTACCACCC	CGCCTTGACG	ATTACCCGAT	1860
	GAGGGTAGTT	CGGGTGGCTG	CTATCAAAGA	CACACCCCTC	CGACCCCGGG	ATGAGCGTTC	1920
	CGTGGATCGC	GAATCCCAGC	CCCGTCACCA	GTTTGTGCCAG	GGTCGAAGGG	GGCTTGCACC	1980
	CGCGGCTGAT	GGCCCGAAGT	GCCTCCCGGT	CCATGGTGTC	GAGCTCTTCC	GGGGTGAACC	2040
15	CCGTGGCCCC	CATCTTACCG	CTGTGCGAGG	TCCGGACGTC	GGGGGGGGCT	GCGCGGCCCC	2100
	GAACAGGAGG	ATGGCCGCTG	GCTCCGGGCA	GGGGGGCGGC	CGAAACCATG	GACAGAAAAC	2160
	GCCCCCTCCG	GTAGTCCTCC	GGGTAGGCCA	CGTCATCCGG	GGCGTCATCG	TCGGCCTCGT	2220
	CTTCCTCCTC	CGCACCCGCG	GCGTCCACGA	TGGGGTAGTC	CTCGTCGCTG	TGCATCTGGG	2280
	CCAAGATCTC	CTGCAGCTGA	CACAGGCGCG	CAGCCTCGCC	GGGGGACGGT	GGGCGGGAAG	2340
20	GGTGGATGGT	TTCCGGGGGC	CCGGGGGCCA	GGTACGCATC	CTCCGCGGGG	GTATAAAAGG	2400
	TGCTCGCCCG	GAAGGCCGGG	GCCGTGTTTG	TCTCCGGCGG	GACGGACGCC	TCCTGTCTCT	2460
	TGTCGGGTCT	ACGGTAGACC	CCACAGAACT	TACGACAGGC	CATTGCGCCG	GTCGCGCGTG	2520
	CCAACCAACG	AGCACCCCGA	GCGACGGGCC	CCGGTGTTTT	AAGAAGCGGC	AGTTTGTCTGA	2580
	CACACCCCCC	CACTACCCCC	GCCCCCTATA	TCCGGAACGT	CAGATTATCC	GGGATACCTA	2640
25	GCCAACCAAA	CAAGGCTGAA	AAAATCGAAC	GTGCGAACGG	GCCGTGTGAT	AGCAAGCAGC	2700
	CCCCCGGGGT	CCGCGCGCCG	TCCCCGCCGT	CATAGGTCCG	CAGACAGGCG	AGTGAGTGAA	2760
	GATCGGACCA	CGGGCCTAAT	ATACCGACAT	GGGCGTTGTT	GTGTAAGTG	TGGTTACCCT	2820
	CCTAAACCAA	CGAAACGCCC	TGCCGCGGAC	TTCGCTGAC	GCAAGCCCGG	CTCTGTGGAG	2880
	TTTTCTGCTT	CGGCAATGCC	GGATCTTGCC	CTCCGAGCCT	CTGGGAACCC	CGGTGGTGGT	2940
30	TCGCCCCGCG	AACCTTCGCA	GGCTGGCCGA	GCCTCTGATG	GACTTGCCCA	AATTACCCCG	3000
	ACCGATCGTG	CGAACCCGCT	CCTGTCGCTG	TCCCCCAAAC	ACCACGACGG	GCCTGTTTGC	3060
	GGAGGACGAC	CCCCTGGAAA	GCATCGAGAT	TCTGGATGCC	CCTGCGTGTT	TTCGGCTCCT	3120
	GCATCAAGAG	CGCCCCGGCC	CCCACCGGCT	ATACCACCTG	TGGGTGGTCG	GGGCGGCGGA	3180
	CCTGTGTGTG	CCGTTTTTTAG	AGTACGCACA	AAAAACCCGG	CTGGGGTTTC	GCTTCATCGC	3240
35	CATGAAGACC	AACGACGCGT	GGGTGGGGGA	ACCGTGGCCC	CTGCCCAGTC	GGTTTTTGCC	3300
	CGAGCGGACC	GTGTCGTGGA	CCCCGTTCCT	CGCAGCGCCT	AATCACCCCC	TGGGAAAATC	3360
	TCCTTAGCCG	ATACGAATAC	CAATACGGCG	TGGTGGTGCC	CGGCGACCGG	GAACGCAGCT	3420
	GTCTTCGCTG	GCTACGGTCC	CTCGTGGCTC	CTCACAACAA	ACCCCGCCCC	GCATCATCCC	3480
	GCCCCGATCC	GGCGACCCAC	CCCACGCAGC	GCCCGTGTTT	TACGTGCATG	GGGCGACCCG	3540
40	AGATTCCCCG	TGAGCCCTCC	TGGCAGACGG	GGGACGATGA	CCCCCAGAAC	CCCGGGCCCC	3600
	CGCTGGCCGT	TGGCGACGAG	TGGCCTCCGT	CATCCACAGT	TTGCTATCCA	ATCACCAACC	3660
	TCTAACCCCC	CCCCGATGCT	AATAAAAAAC	ACTGCGCCCC	ATTACACGTA	CGAGCGGTGT	3720
	CGCGTTTGTT	TCTTTTTTTG	TCGTTCCCTC	CTCCACCCCC	AGAAAAACCA	GACACTCAGA	3780

	CACAAAAGCT	TTCTTGTAGG	GCGTTTATTT	TCGTTTGGCA	AACACACCGG	GCTGGGGGTC	3840
	CCGCGCTCAT	GGCCGGAGAA	ATGTGTGGCC	GCAGGGATAG	GGGCAGGCGG	CGGGGAAGCG	3900
	CATTTTTCGG	CACCCGTCCT	CGCGTTCTGG	GGCTTCCGTT	GCGCGACACG	CGCCCGGGGC	3960
	GGGTAGGCCA	GAGCGTTCGG	CGGAGCTGGT	ATCGGCGACC	ACCGCGGACA	GCCAGGGCTG	4020
5	GGAGCCCTCC	TCGGACGTCC	AGTCAAACGT	CCCAAACCTCA	CACTCCAGCC	GAGCCGCGAT	4080
	AGCCCCGCCG	GACGCGGATT	CCGGGTTCCT	CCGGCCGGCC	GGGGAGGGCC	CCCCCTCCGT	4140
	GTCGGAATGG	GACGCGAGGG	TATCGTCCGA	GTCCGTTCGA	TCCGATATCT	CATCATCGCT	4200
	CGTCGAGCCG	CCGTTGGTCT	CGAGATTGCC	AACATCACAC	TCGGGGCCGT	AACGCTTGGC	4260
	GTTTCAGACAG	GGCAGCTCGC	ACACGGGCTC	GGCGGCGGGT	TCAAAACGCG	CCTCGACCTC	4320
10	CCGGATTGCG	TTTCGCAGGC	GCACGTCCCA	GGTTCCGCCC	GAGATCTGCA	GCAGGCGGCC	4380
	CCACGTGCGC	GGCCCCAGGC	GGGTCCGGCA	GTAGCCCATTA	AGGTAGCAGT	CTCGCACCAG	4440
	GTGGCGCAGG	CGGTTGGCGC	TGCCGGGCGG	GTTCCGGGCG	GCACGCAGGA	TCCGAAAGAG	4500
	CTGGTTGACG	CACTGGCGTA	TGTAATGCAG	ATCAAGCGTC	CAGGCGGCTC	CGCTCCGCAT	4560
	CTGGGCCCTGG	CGAGCAGAGC	GCCGCGGTGC	GGCTGCGCGA	TCCCCGGAAG	ACTGGGCCGG	4620
15	GGCCTGGGGT	TGCGCCGCGC	GGATAGGTCT	GTCGTTTCTC	CACACCTCGG	GGAAAACCAC	4680
	ACCCGCGCGC	CGGTCCGGGG	AGCTCGTTAA	TCGCAGGTTG	ATTCGGGGTC	GCTTGGATTT	4740
	ACGGGGGGGG	GTGTCAACCA	ACCAGCCGTC	TGACGCATCA	TCATCATCGT	CGTCCGACGG	4800
	CCCCGTTCCT	TCCGATTCCG	TCCCCGTGGT	CGATTCTGCC	GACAGATCCA	AAAAATACCT	4860
	CCCCCGGAGC	TCCCGCGGGG	AGCGACGGCG	CCCGCCGCGT	AGGTCTCCCG	CCTCATCCTC	4920
20	GGACGACTCG	GTCGAGGAGG	ATTCCGATTC	TGTGTCGGGC	TTACCCCTCAG	ATTCGACGCA	4980
	GCTGGGGAGG	ACGGGGCGTC	TGCGCTTCCG	TGAACCCGGG	GGTGGGGATG	GGGGAGCATG	5040
	ATTCGCAGGC	GTCGTGTTGA	CCGCGGGCGG	GTCGGGGGGG	ATGTCTGCCA	TAGTGGCGAC	5100
	GCCTTGTCGC	CAGTTACCAC	ACCGGTGTCC	CGTCCACGAA	GGCGGCGCCC	GGCCTGCGAT	5160
	AAAGCGCGGA	TGTTGGGATC	GGGGCCCCCC	CCCCCGTCT	CCCTTTTCCC	CTCTCTTCTC	5220
25	TCTCCCTCCT	CCCTCCCCCC	CTCCTCTGTC	TCTCTCCCT	TTTTCCTCA	CCTCCCCCTC	5280
	TTCTCCTCCC	TCCTCCTCCT	TCCCTTTCCC	CTCCCCCTCC	TCTCTCTTCC	TCCTTCTCCT	5340
	CCCCTCTTTC	TTCTTCTTCC	CTCTCTCCCT	TTCCCCCACT	CTCATCTTTC	CCACTTCGCT	5400
	CCCTTTTCTC	TCTCTTCCCT	TCTTTTCCTT	TTCCCTACCC	TCTCCCTTCT	TCTTCCGTCT	5460
	CCCCTCCCCCT	TCTCTCCTCT	CTCTCTCCTC	GTCCTTGTAT	CCACGCTACC	TCCTCTTCAT	5520
30	CTCATCTCTT	TCTTCTCTCT	TTCTCCTCCC	TCTCCCTCT	TTTATCTCCC	CATTCTTAC	5580
	TCTCTCCTTA	TCTCTACCTT	TATCTCAACA	GCTCTCTCAC	GCTCCTCCCA	TGGCCATCTC	5640
	CTCTCCTTCC	TCTCCCCCTC	TCTCACATCT	CATCCTCTCT	TCTCCTTCCA	TCTCATCTCC	5700
	CATCTCCTGC	CCCCACGCTT	CTCCCCCTCT	CTCTTCTCAC	T		5742

35

(2) INFORMATION FOR SEQ ID NO:94:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 507 amino acids

(B) TYPE: amino acid

40

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

```

Val Gly Gly Cys Val Asp Lys Leu Pro Leu Leu Lys Thr Pro Gly Pro
5   1           5           10           15
Val Arg Ala Arg Trp Leu Ala Arg Ala Thr Arg Arg Met Ala Cys Arg
      20           25           30
Lys Phe Cys Gly Val Tyr Arg Arg Pro Asp Lys Arg Gln Glu Ala Ser
      35           40           45
10  Val Pro Pro Glu Thr Asn Thr Ala Pro Ala Phe Pro Ala Ser Thr Phe
      50           55           60
Tyr Thr Pro Ala Glu Asp Ala Tyr Leu Ala Pro Gly Pro Pro Glu Thr
65           70           75           80
Ile His Pro Ser Arg Pro Pro Ser Pro Gly Glu Ala Ala Arg Leu Cys
15           85           90           95
Gln Leu Gln Glu Ile Leu Ala Gln Met His Ser Asp Glu Asp Tyr Pro
      100           105           110
Ile Val Asp Ala Ala Gly Ala Glu Glu Glu Asp Glu Ala Asp Asp Asp
      115           120           125
20  Ala Pro Asp Asp Val Ala Tyr Pro Glu Asp Tyr Ala Glu Gly Arg Phe
      130           135           140
Leu Ser Met Val Ser Ala Ala Pro Leu Pro Gly Ala Ser Gly His Pro
145           150           155           160
Pro Val Pro Gly Arg Ala Ala Pro Pro Asp Val Arg Thr Cys Asp Ser
25           165           170           175
Gly Lys Met Gly Ala Thr Gly Phe Thr Pro Glu Glu Leu Asp Thr Met
      180           185           190
Asp Arg Glu Ala Leu Arg Ala Ile Ser Arg Gly Cys Lys Pro Pro Ser
      195           200           205
30  Thr Leu Ala Lys Leu Val Thr Gly Leu Gly Phe Ala Ile His Gly Thr
      210           215           220
Leu Ile Pro Gly Ser Glu Gly Cys Val Phe Asp Ser Ser His Pro Asn
225           230           235           240
Tyr Pro His Arg Val Ile Val Lys Ala Gly Trp Tyr Ala Ser Thr Asn
35           245           250           255
His Glu Ala Arg Leu Leu Arg Arg Leu Asn His Pro Ala Ile Leu Pro
      260           265           270
Leu Leu Asp Leu His Val Val Ser Gly Val Thr Cys Leu Val Leu Pro
      275           280           285
40  Lys Tyr His Cys Asp Leu Tyr Thr Tyr Leu Ser Lys Arg Pro Ser Pro
      290           295           300
Leu Gly His Leu Gln Ile Thr Ala Val Ser Arg Gln Leu Leu Ser Ala
305           310           315           320

```

	Ile	Asp	Tyr	Val	His	Cys	Glu	Gly	Ile	Ile	His	Arg	Asp	Ile	Lys	Thr
				325						330					335	
	Glu	Asn	Ile	Leu	Ile	Asn	Thr	Pro	Glu	Asn	Ile	Cys	Leu	Gly	Asp	Phe
				340					345					350		
5	Gly	Ala	Ala	Cys	Phe	Val	Arg	Gly	Cys	Arg	Ser	Ser	Pro	Phe	His	Tyr
				355					360					365		
	Gly	Ile	Ala	Gly	Thr	Ile	Asp	Thr	Asn	Ala	Pro	Glu	Val	Leu	Ala	Gly
				370				375					380			
	Asp	Pro	Tyr	Thr	Gln	Val	Ile	Asp	Ile	Trp	Ser	Ala	Gly	Leu	Val	Ile
10																
	385					390					395					400
	Phe	Glu	Thr	Ala	Val	His	Thr	Ala	Ser	Leu	Phe	Ser	Ala	Pro	Arg	Asp
						405				410					415	
	Pro	Glu	Arg	Arg	Pro	Cys	Asp	Asn	Gln	Ile	Ala	Arg	Ile	Ile	Arg	Gln
						420				425					430	
15	Ala	Gln	Val	His	Val	Asp	Glu	Phe	Pro	Thr	His	Ala	Glu	Ser	Arg	Leu
						435				440					445	
	Thr	Ala	His	Tyr	Arg	Ser	Arg	Ala	Ala	Gly	Asn	Asn	Arg	Pro	Ala	Trp
						450				455			460			
	Trp	Ala	Trp	Thr	Arg	Tyr	Tyr	Lys	Ile	His	Thr	Asp	Val	Glu	Tyr	Leu
20						465					475					480
	Ile	Cys	Lys	Ala	Leu	Thr	Phe	Asp	Ala	Ala	Leu	Arg	Pro	Ser	Ala	Ala
						485					490					495
	Glu	Leu	Leu	Arg	Leu	Pro	Leu	Phe	His	Pro	Lys					
						500				505						

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 188 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

	Met	Gly	Val	Val	Val	Val	Ser	Val	Val	Thr	Leu	Leu	Asn	Gln	Arg	Asn
	1				5					10					15	
40	Ala	Leu	Pro	Arg	Thr	Ser	Ala	Asp	Asp	Ala	Leu	Trp	Ser	Phe	Leu	Leu
				20					25					30		
	Arg	Gln	Cys	Arg	Ile	Leu	Ala	Ser	Glu	Pro	Leu	Gly	Thr	Pro	Val	Val
			35					40					45			

Val Arg Pro Ala Asn Leu Arg Arg Leu Ala Glu Pro Leu Met Asp Leu  
 50 55 60  
 Pro Lys Phe Trp Ile Val Arg Thr Arg Ser Cys Arg Cys Pro Pro Asn  
 65 70 75 80  
 5 Thr Thr Thr Gly Leu Phe Ala Glu Asp Asp Pro Leu Glu Ser Ile Glu  
 85 90 95  
 Ile Leu Asp Ala Pro Ala Cys Phe Arg Leu Leu His Gln Glu Arg Pro  
 100 105 110  
 Gly Pro His Arg Leu Tyr His Leu Trp Val Val Gly Ala Ala Asp Leu  
 10 115 120 125  
 Cys Val Pro Phe Leu Glu Tyr Ala Gln Lys Thr Arg Leu Gly Phe Arg  
 130 135 140  
 Phe Ile Ala Met Lys Thr Asn Asp Ala Trp Val Gly Glu Pro Trp Pro  
 145 150 155 160  
 15 Leu Pro Asp Arg Phe Leu Pro Glu Arg Thr Val Ser Trp Thr Pro Phe  
 165 170 175  
 Pro Ala Ala Pro Asn His Pro Leu Gly Lys Ser Pro  
 180 185

20 (2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 amino acids

(B) TYPE: amino acid

25 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

Met Gly Arg Pro Glu Ile Pro Asp Glu Pro Ser Trp Gln Thr Gly Asp  
 1 5 10 15  
 Asp Asp Pro Gln Asn Pro Gly Pro Pro Leu Ala Val Gly Asp Glu Trp  
 35 20 25 30  
 Pro Pro Ser Ser His Val Cys Tyr Pro Ile Thr Asn Leu  
 35 40 45

40 (2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 515 amino acids

(B) TYPE: amino acid



(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

	Val	Gly	Arg	Met	Arg	Val	Gly	Glu	Arg	Glu	Arg	Gly	Lys	Lys	Lys	Lys
	1				5					10					15	
10	Glu	Gly	Arg	Arg	Arg	Arg	Lys	Arg	Glu	Gly	Gly	Glu	Gly	Lys	Gly	Lys
					20					25					30	
	Glu	Glu	Glu	Gly	Gly	Glu	Glu	Gly	Glu	Val	Arg	Glu	Lys	Gly	Glu	Arg
					35					40					45	
	Asp	Arg	Gly	Gly	Gly	Glu	Gly	Gly	Gly	Arg	Glu	Lys	Arg	Gly	Glu	Lys
15					50					55					60	
	Gly	Asp	Gly	Gly	Gly	Gly	Pro	Arg	Ser	Gln	His	Pro	Arg	Phe	Ile	Ala
	65						70				75					80
	Gly	Arg	Ala	Pro	Pro	Ser	Trp	Thr	Gly	His	Arg	Cys	Gly	Asn	Trp	Arg
							85				90					95
20	Gln	Gly	Val	Ala	Thr	Met	Ala	Asp	Ile	Pro	Pro	Asp	Pro	Pro	Ala	Val
							100				105				110	
	Asn	Thr	Thr	Pro	Ala	Asn	His	Ala	Pro	Pro	Ser	Pro	Pro	Pro	Gly	Ser
							115				120				125	
	Arg	Lys	Arg	Arg	Arg	Pro	Val	Leu	Pro	Ser	Ser	Ser	Glu	Ser	Glu	Gly
25							130				135				140	
	Lys	Pro	Asp	Thr	Glu	Ser	Glu	Ser	Ser	Ser	Thr	Glu	Ser	Ser	Glu	Asp
	145						150				155				160	
	Glu	Ala	Gly	Asp	Leu	Arg	Gly	Gly	Arg	Arg	Arg	Ser	Pro	Arg	Glu	Leu
							165				170				175	
30	Gly	Gly	Arg	Tyr	Phe	Leu	Asp	Leu	Ser	Ala	Glu	Ser	Thr	Thr	Gly	Thr
							180				185				190	
	Glu	Ser	Glu	Gly	Thr	Gly	Pro	Ser	Asp	Asp	Asp	Asp	Asp	Asp	Ala	Ser
							195				200				205	
	Asp	Gly	Trp	Leu	Val	Asp	Thr	Pro	Pro	Arg	Lys	Ser	Lys	Arg	Pro	Arg
35							210				215				220	
	Ile	Asn	Leu	Arg	Leu	Thr	Ser	Ser	Pro	Asp	Arg	Arg	Ala	Gly	Val	Val
	225						230				235				240	
	Phe	Pro	Glu	Val	Trp	Arg	Asn	Asp	Arg	Pro	Ile	Arg	Ala	Ala	Gln	Pro
							245				250				255	
40	Gln	Ala	Pro	Ala	Gln	Ser	Ser	Gly	Asp	Arg	Ala	Ala	Ala	Pro	Arg	Arg
							260				265				270	
	Ser	Ala	Arg	Gln	Ala	Gln	Met	Arg	Ser	Gly	Ala	Ala	Trp	Thr	Leu	Asp
							275				280				285	

Leu His Tyr Ile Arg Gln Cys Val Asn Gln Leu Phe Arg Ile Leu Arg  
 290 295 300  
 Ala Ala Pro Asn Pro Pro Gly Ser Ala Asn Arg Leu Arg His Leu Val  
 305 310 315 320  
 5 Arg Asp Cys Tyr Leu Met Gly Tyr Cys Arg Thr Arg Leu Gly Pro Arg  
 325 330 335  
 Thr Trp Gly Arg Leu Leu Gln Ile Ser Gly Gly Thr Trp Asp Val Arg  
 340 345 350  
 Leu Arg Asn Ala Ile Arg Glu Val Glu Ala Arg Phe Glu Pro Ala Ala  
 10 355 360 365  
 Glu Pro Val Cys Glu Leu Pro Cys Leu Asn Ala Arg Arg Tyr Gly Pro  
 370 375 380  
 Glu Cys Asp Val Gly Asn Leu Glu Thr Asn Gly Gly Ser Thr Ser Asp  
 385 390 395 400  
 15 Asp Glu Ile Ser Asp Ala Thr Asp Ser Asp Asp Thr Leu Ala Ser His  
 405 410 415  
 Ser Asp Thr Glu Gly Gly Pro Ser Pro Ala Gly Arg Glu Asn Pro Glu  
 420 425 430  
 Ser Ala Ser Gly Gly Ala Ile Ala Ala Arg Leu Glu Cys Glu Phe Gly  
 20 435 440 445  
 Thr Phe Asp Trp Thr Ser Glu Glu Gly Ser Gln Pro Trp Leu Ser Ala  
 450 455 460  
 Val Val Ala Asp Thr Ser Ser Ala Glu Arg Ser Gly Leu Pro Ala Pro  
 465 470 475 480  
 25 Gly Ala Cys Arg Ala Thr Glu Ala Pro Glu Arg Glu Asp Gly Cys Arg  
 485 490 495  
 Lys Met Arg Phe Pro Ala Ala Cys Pro Tyr Pro Cys Gly His Thr Phe  
 500 505 510  
 Leu Arg Pro  
 30 515

## (2) INFORMATION FOR SEQ ID NO:98:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 6328 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

TGCGGTCGAC TCTAGAAGAC CCTGTGCACG GGA CTGGTT GGGCGACGTC TGCGGTCTAN

60

	TGGTCGGCGT	GGGGACCGGC	TGTGTGGTGG	GTGGGGGAAG	CACGTTGGAC	TACAACCGAC	120
	CACACGACAT	GCAGGCTTCT	GAGCCGCGAA	TCCCCACCAT	GGCGTTGGGG	GCTGGGCATG	180
	CCCACGCATG	CAGGGATGAC	GGAGATGATA	GCGTGATTGA	CGCCCCGCCC	CCATACGAAA	240
	TGTGCGCCGG	CGCGAGCGCG	GGCCANTTTG	TCGTTATTGA	TATCGACACC	CCCACGGACT	300
5	CGCCTCCACC	GTACTCTGCA	GGGACGTCTC	CCGTTGGGCT	TGTTTCACCG	GCTTCTTCCG	360
	GTGACGGCGA	GGTGTGTGAG	CGTGGCCGCT	CGCGCCGCGC	CGCCTGGCGG	GCCGCTCGGC	420
	GCGCCAGGCG	CCGCGCCGAA	CGACGGGCGC	GGCGCCGGAG	CTTTGGCCCA	GGGGGGTTGT	480
	TTGTGGAGAC	CCCCCTGTTT	CTACCGGAAA	CTATGATTGG	GGCCACCCCT	GGCGTGGGAG	540
	GCGACCTCCC	GTCGGGCCTC	CCTACTTACG	CAAAGGCGAC	CTCGGATCGC	CCCCCACCT	600
10	ACGCCATGGT	CATGGCCGCA	TGTCGACCG	AGCCACCGGG	CGGGTCCGTG	GGGCGGCGCG	660
	ACCAACCCCG	CGTGCAAAGC	TCGCGCACGT	GGCGACCCCC	GCTCGTCAAT	TCGCGAGAGC	720
	TGTACCGGGC	CCAACGCGCG	GCCCCTGTG	CGTCAAGCTC	CGACACGCCC	CAAGCCCCAG	780
	GGTGGTGTGG	CGGGACGTGT	CGTCATGCGG	TTTTTGGGGT	GGTCGCGGTG	GTCGTCGTTA	840
	TCATCTTGCG	CTTCCTTTGG	CGGTAAGCTT	CCCCCTCCC	GCGATACAA	GAATAAAAGT	900
15	CGCGTTAACA	CACACGCTGG	TTCGTCGCGT	GGTATTTACC	GGGTTCTTAT	AACCCACAAA	960
	CTCACACCGC	GTCTGTTTTG	GTGGTTCTC	ACTCTTTATT	AATGAGGTTG	CATACGGACT	1020
	CGGAGGGAAG	GGGGTGGGTT	ATACCTTGAT	TTTGATTTTG	ATTTTGTGGC	GTCGCTGTTT	1080
	TGCCGCGCGA	GCGGCCGTGC	CGCTTGAGCT	TATAAAGCGA	AGGTGTTGTA	GGGCCGCGGA	1140
	TGCCCCGAGC	CAGCAGGTTT	TGGAGAACGG	ATACCGACAG	TGACAGTGGT	ACACGATACC	1200
20	GTTTATCGTG	TATTCCCCCG	CGGGGGATGC	ACCGGAGACG	TCCTTGATGG	TTCCGCACT	1260
	GAATGGGCCC	GTGCACCGCA	GCCGCCCCCC	ACGCTTATCC	TCTAGTTCGC	GCAAGACGGG	1320
	CGGGTGGTTG	ACCAGGTCCG	CGAAGTTGCG	GAGCTCGGTA	ACCAGCGGAA	GGGAGTGCTG	1380
	CGAGTCCTTG	TACACCGCAA	AGAAAAACA	GTGGATGGGG	CCGCTGGTCG	AGCAGGAGGC	1440
	GCGGACTAAA	TAACCTCCGC	CGGCCAGGCA	CGCGGGGTCC	GTTTTGGTCG	TGTGCAGGGC	1500
25	GTTTTCGGCC	TGCCACGTGG	CATTGAGACA	GTACGGGGGG	GCGACGTGGG	TGATGTCCGG	1560
	GGCCCGTAAA	AACAGGTTG	AGAGGGGCGT	CGTTGTCATG	TTGCGAGGGG	GGGGGTGCGT	1620
	GAATACGCGT	CTGCCGTGAG	TTCTTGAGCG	CGCTATGAAG	CGGGCCGGGC	CGGGGCCCCA	1680
	CATTTATCCG	GTGGGTCATC	GCCCTCCTCC	CACGCGCACG	CCGGCATCGC	CCCGAGTCT	1740
	CCGCCCCACC	CGCCGCGCGC	GCCAAGAACA	TCACACGGAA	CCACTTGGGT	TGACGTCAAT	1800
30	ATGTTTATTC	TTGCCATAAA	TAGGGAGTTG	CAGTAGAAGT	ATTTGCCGTG	CACATATAAG	1860
	GGGGCGATAG	TGTGACTGGC	CGTCAGCTCG	CACACGCGAC	TGGAACACTC	CTGGCGGTGC	1920
	GTGTCCAGTA	TTTCAATGAG	ACCCGCCATG	CAGGCCCCCG	GGATGTAAAA	GTGCATCGTC	1980
	TCGCCGGCCC	CAACCCCCAC	GGTCGTGTAG	TCGATCTCCG	ACACGCCGCG	CTCGACGCGG	2040
	TTGGCGAGGC	GGGCCAGGAT	GACCAACACA	AAGGAGGCAA	TATCCTTAAT	GTCCGACAGG	2100
35	CGTCGCCGCG	AGCACAGGTC	GTCCAGCCCG	CACAGGCCTC	GGGCCTTCAG	GTAGCACTGC	2160
	AGAAAGGGGC	GCAGGCGCGT	GGCGAGGTTT	TCCAGCACGG	CGGCCGCCGT	TCCGATGATA	2220
	GGGTCCTGGG	GGCGGAGCGG	CAGATTGTGG	TGAATGCACA	TCTTGACACA	CGCCAGCGTC	2280
	TCATCCGCGG	ACGCCAGGGC	CTCGATGAGA	TTTTCTTGGC	GCAGCACGCA	GTCGCGCATG	2340
	GCCTTGGCTG	TCGACGCGGC	CCGCGGGTTG	GCTGCGAATG	TGCGGTAGAG	GCTCGGGCCG	2400
40	TGAGCGACCA	GGGTTTCCCA	GGAAACCCGA	CGGGTCTCGG	CGTCAAACCC	CCCCGCTTGG	2460
	GTGGCCAGCA	CGGGAGCCCA	GGGGCTGTTC	GCGGCGGGAA	ACGGCATCCC	GCCAAAGGGG	2520
	TCTTGCAATG	CCAGGGCACT	GCGTCCAAAG	CTTTCGCTGA	TGCGCTCGAC	CGCCGCGCGC	2580
	TCGGATATGG	ATCGCAGAAC	CGCCCGAACG	GCGGGGTCTGA	TGGTGTGCGG	AGAGGGCGCC	2640

	TTTCGCTCCG	GGACCGGGG	GCGGCCGTCC	GCGTGCGGG	GGGTCAGGA	CAGCGCCATC	2700
	AGCGGAGGG	GGGCTGGCG	CGTGCCCTCGT	GGCCGCGGG	CCCCCGAGA	CGTCGAGCTG	2760
	CTCCCTCCG	CGCCGCGCCT	CGCCGTGGGT	GGCGCCGGG	CCGTCCGTCC	GCGCCGACGC	2820
	GGGGTGGCAA	CCCCCTTGGT	TGTGGGCGTT	TCTGGAGACG	CGCCGCGGG	GGTTTGGTGT	2880
5	GGAGTCGGCG	CCGCCGGGG	CGTATTGACC	CCGGCCCCGG	CGGCGACCTC	GCCGCCCGCC	2940
	TCGGGGATGC	GGTGCCCTGG	TCGACGGGG	TTGGATGCGG	GCCACCTTCC	CCCCGTGCGG	3000
	TTCCCGGGG	GAAGCCGACC	GCCTGGTCCC	GAGGCGCGAC	CACACGCCGG	TGGTCGCGGG	3060
	TGGCGGATCG	TCGGCTCCCC	GCCGCGCTGC	CGGGCGAGGC	GTCAAGGCTT	CGGGGTGCC	3120
	GGCGTCCTCG	GGGCGGGCCG	GGGACCTTT	GGGAATCGCC	GCGTAGATGG	CCTCCGCCCC	3180
10	TCCGTCTCCG	CAGGGGTCTT	CCATGTCCCT	GTCCGACAAG	GAACACTCCC	CGTGCTGTG	3240
	GGACTCGGG	TCGTGCGGG	GGCCCTCCTC	GTCCCGCTCC	AGAGCGTCCT	CCTCGAGCTC	3300
	GCTGTGCGAC	AGGTCCAATC	CTAGGTCGAT	TAGCATATCA	ATGTGCGTAG	CCATGTTGTA	3360
	GGTCGCCGG	GCTGGGATGG	CGGGTGTCTT	CCGAGGGGG	GCGTGTGCGA	AGAGAGTGGC	3420
	CGGGTCCGAA	TCGAGGAGCG	GCACCGACGC	GCAACCGGG	TCGGCACACG	GCAGCACACA	3480
15	GCGCCAGTG	GGCCGGTGGC	TGCCCTTATA	CCCGCACGAC	CGGGGCCGGC	TTTCCGAAAC	3540
	TCCTCCTTGT	CCCTCCCCGT	CGGGCGTCAC	CGCCCCCGCC	CCCGCCGTCC	CCAGAAACCA	3600
	ATCGGACGCC	GAGGGTGGGT	TTTATGTATT	TAATTAGCAT	ACGGCAGGTC	TGGGTCCGCC	3660
	TTCGCGTACA	CGCGTAGGCG	GGGTGCGGA	AGCACGCGGT	AGGGTGGGGT	GTATGCGGAA	3720
	GTGCGACGAG	CCTGCCCTGTG	CTGGACCGGG	GGAGGGGCAA	GCAGACCCGA	GGCCGGATCG	3780
20	GCTCTGTGCA	CGATTTTAAT	TTGCATGCGA	CGTGCGAGGG	TGCGTAGGCC	CGAGCGGGT	3840
	CGTGTATTTA	ATTTGCATGG	GCGACTGGGT	CCGCCTCTTC	CAACGGAAGA	GGCGTTACGT	3900
	CACAGATCAA	ACAGGCGCCG	CTGAATCTCC	TGTTTCGTAGC	GAAGCGCCAT	AAGCACCACC	3960
	CCGGCCACGA	CGGCGATATA	GCACAGGCGC	ACGGCGATAC	CGGAGAGGAT	GATGGAACAG	4020
	CAGCGCCCGC	AGACGCCCGA	CCACCCTTTG	GAGCGCCCC	TGGGGGCCGC	TGGTTCCGCG	4080
25	TTTTTGGGG	CCGAGTCCCG	CCGCAGGATG	AAATACAGCT	CCGTCAGGGC	GATGATGGAC	4140
	ACGAAACACC	AGGTGGTGAT	TGTTAGAAAC	AGGGGGTATG	TGATCGCGCA	GGCGCCCCGG	4200
	GAGATGAGAG	CGGTGCCGAC	GATGAGACCG	AGGGCCACGA	AGCGGAGCAG	CAGCTCGCAG	4260
	CCCACGATGA	CGCCAACGGC	CGGGCGGTGG	TACAAGAAGG	TGACCGGATC	CGCCTCGAAC	4320
	AGCTGCACCA	GGGTCTGGCG	TTGAACGGAT	AGCTCGCAGA	GGAGGCGGGT	GATTTTCGTG	4380
30	TAGGGGTATT	GCAAGAACAC	GCTCGACACT	ATGCGGCCCG	CGTAGTTCAA	AAGATAGGTC	4440
	GCCGGGGCCA	CCATGTCTGTG	CGCGGGAATC	ACGACGCCGA	ACATGCATCG	TCGTTGGTGA	4500
	AGGGCGACGA	ACGCTAGATA	CAGAAACCAA	CCGACGACCA	CCAGGCGCAT	CTGGGTGTCC	4560
	CAGAGGGCCT	CCAAGCAGTT	TACGGCCTCG	TGCACGTTCA	TGACCCGGCG	GCTCATGGCG	4620
	CCGGGGATGG	CCGGGAGGGA	CACGGCCCGA	CCTTCGATGA	TATTGGCGTA	GCAGACGTGG	4680
35	GCGTGGGGGG	TCCATGCCCC	GCCGGGGGGG	GCGGTGCGCG	GGCCAGAAA	CAACAGCGTC	4740
	TGGTTTATCT	TCATCCACAC	GAGGGCGGTA	TCGTTGTGTG	CCCCGCGGG	GCGCACC CG	4800
	TAAATACATC	GGTGGAGCGG	ACTGGCACCA	AAGACGATGT	ACCACGCGAG	CACGAGGCCG	4860
	TAGGCCGTTA	TGAAGATGAC	GGTCGTGAGG	TGCTGGAGGG	AGCGGACCGC	GAGCATGGCG	4920
	TGCCCCGATC	GACGGTAAAC	AGCGTGTGCA	GGCGGTTGTT	ATCGCATTTA	GTGGCAAAGC	4980
40	ACTGCTGACA	CAGGGACGCG	CATAGGCGGT	TGTTGGTCCC	GACGCTCAGC	GCGAGAAAGG	5040
	TCCGGGCCGT	CGCGCGACTT	GCCCTGCCGT	GCTTGAAGCG	CAGACACGAG	AGGCTTTGGT	5100
	TCAGGGCGCC	GCGGCCGGGG	ATCAGCTGCA	GCAGGACCCA	GTCGTCCTTA	ATGACGGCCC	5160
	GGCGAACGGA	CACGGCCTGA	TATTCGCCCG	CGTGCTCGGG	AAAGTGGGCC	TCGATGCACG	5220

CGACTATTCG CCCAGCAGT TCTGACGCGA ACCCTCCAC GGTCTCCCCG GCGTCGGGCC 5280  
 GCACATCGTA GCGGCCAGG ACCTCCGTCA GGGTCTCGCG TCGCCCAAAG TGCTCCAGGG 5340  
 CGTTGCGCGA CGCCTTCTTC TCGAAAAAGC TGACATAGTC CCCGCCAGG CTGTGGAGGA 5400  
 CGCGGATCTC CCGCGGGGCC GCGGAAAACA TGGGCGGGG GTGAAAGTGG AAGCGCCGCG 5460  
 5 GGTGCGCGTG CGCGGCGACG AACGCCGGA CGTCCTCGCA CGCGGGGGG ATCACGAAGA 5520  
 CGGGCAATAA CCGGCCGCAC GCGGAGCCGT CCGGGCCGAT TTTGGCGAAA TACGGCAAGC 5580  
 GCAGGCTGTG GCCGTGGGCG TACACGCCCG TATCGATCAG CAAAAAGTTC TTTACGTGGC 5640  
 TCCCTACGGC CTCCACGAAG TTGCGGTCCA ACAGACCGC CTGCTGGATC ACCCTCGCCA 5700  
 CCCACGCAT TGTCAGGGAG CCGTGACAAA CGTACGGGGC GGGGACCGGT AGGCACACGC 5760  
 10 GCAGCCCGAT CTTGTGCGCG CAGGAACACA CGACGGAGTC CGGTTCGCTG GCGCTCGCCG 5820  
 CTGTATCTG TTCGTGTAGC AGGTCGAGGT ACGCGCCTC GTCGTCCGGG AGGGGGCCGT 5880  
 GGGTCGTGTC CATGGGGTCC GTGTCTCCT CCCACTCCTC GTCGCCGTCG TCGCCACCGG 5940  
 CGTCGGGGAA CCAGTCCCCG TCGCCGTCGT CGCCACCGGC CGAGGGCCCG TCGCCCGCAC 6000  
 AGACGGGCGG CGCGCGGGG CGACAGGCGC TTTTGAAAAA ATAACAGGGA TAGGCGTCGG 6060  
 15 GGTTTACGCG GGCCGCGGGA AACAACAGNT GAACGCCGC CAGCGCCCCG CGCATAAAGT 6120  
 GACCCAGGGC CTCGTGGAGC CGGGGAAAGG GGACGGGCTC CTTCAGGGCG ATGTCCAGAT 6180  
 CCAGGATGAT GTTCGTAACG GCCAGCGCGG CGTTGAAGAT CTCGTTGCGG TTCACGTACA 6240  
 TATGCCCCC GACGGCCAGG CCGCCCGGGG GGAGCGCGG CCCC GGCGCA AAAGGGCGGT 6300  
 GACCGGGGAC TTCTAGATCG AATGCAG 6328  
 20

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 86 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

Met Val Met Ala Ala Cys Pro Thr Glu Pro Pro Gly Gly Ser Val Gly  
 1 5 10 15  
 35 Pro Ala Asp Gln Pro Arg Val Gln Ser Ser Arg Thr Trp Arg Pro Pro  
 20 25 30  
 Leu Val Asn Ser Arg Glu Leu Tyr Arg Ala Gln Arg Ala Ala Arg Cys  
 35 40 45  
 Ala Ser Ser Ser Asp Thr Pro Gln Ala Pro Gly Trp Cys Gly Gly Thr  
 40 50 55 60  
 Cys Arg His Ala Val Phe Gly Val Val Ala Val Val Val Val Ile Ile  
 65 70 75 80  
 Leu Ala Phe Leu Trp Arg

85

## (2) INFORMATION FOR SEQ ID NO:100:

## 5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 212 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

```

15 Met Trp Gly Pro Gly Pro Ala Arg Phe Ile Ala Arg Pro Gly Thr His
    1           5           10           15
    Gly Arg Arg Val Phe Thr Asp Pro Pro Pro Arg Asn Met Thr Thr Thr
        20           25           30
    Pro Leu Ser Asn Leu Phe Leu Arg Ala Pro Asp Ile Thr His Val Ala
20           35           40           45
    Pro Pro Tyr Cys Leu Asn Ala Thr Trp Gln Ala Glu Asn Ala Leu His
    50           55           60
    Thr Thr Lys Thr Asp Pro Ala Cys Leu Ala Ala Arg Ser Tyr Leu Val
    65           70           75           80
25 Arg Ala Ser Cys Ser Thr Ser Gly Pro Ile His Cys Phe Phe Phe Ala
        85           90           95
    Val Tyr Lys Asp Ser Gln His Ser Leu Pro Leu Val Thr Glu Leu Arg
        100           105           110
    Asn Phe Ala Asp Leu Val Asn His Pro Pro Val Leu Arg Glu Leu Glu
30           115           120           125
    Asp Lys Arg Gly Gly Arg Leu Arg Cys Thr Gly Pro Phe Ser Cys Gly
    130           135           140
    Thr Ile Lys Asp Val Ser Gly Asp Ala Gly Glu Tyr Thr Ile Asn Gly
    145           150           155           160
35 Ile Val Tyr His Cys His Cys Arg Tyr Pro Phe Ser Lys Thr Cys Trp
        165           170           175
    Leu Gly Ala Ser Ala Ala Leu Gln His Leu Arg Phe Ile Ser Ser Ser
        180           185           190
    Gly Thr Ala Ala Arg Ala Ala Glu Gln Arg Arg His Lys Ile Lys Ile
40           195           200           205
    Lys Ile Lys Val
        210

```

## (2) INFORMATION FOR SEQ ID NO:101:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 286 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

Met Ala Leu Ser Leu Thr Pro Pro His Ala Asp Gly Arg Ala Pro Val  
 1 5 10 15  
 15 Pro Glu Arg Lys Ala Pro Ser Ala Asp Thr Ile Asp Pro Ala Val Arg  
 20 25 30  
 Ala Val Leu Arg Ser Ile Ser Ala Ala Val Glu Arg Ile Ser Glu Ser  
 35 40 45  
 Phe Gly Arg Ser Ala Leu Val Met Gln Asp Pro Phe Gly Gly Met Pro  
 20 50 55 60  
 Phe Pro Ala Ala Asn Ser Pro Trp Ala Pro Val Leu Ala Thr Gln Ala  
 65 70 75 80  
 Gly Gly Phe Asp Ala Glu Thr Arg Arg Val Ser Trp Glu Thr Leu Val  
 85 90 95  
 25 Ala His Gly Pro Ser Leu Tyr Arg Thr Phe Ala Ala Asn Pro Arg Ala  
 100 105 110  
 Ala Ser Thr Ala Lys Ala Met Arg Asp Cys Val Leu Arg Gln Glu Asn  
 115 120 125  
 Leu Ile Glu Ala Ser Ala Asp Glu Thr Leu Ala Trp Cys Lys Met Cys  
 30 130 135 140  
 Ile His His Asn Leu Pro Leu Arg Pro Gln Asp Pro Ile Ile Gly Thr  
 145 150 155 160  
 Ala Ala Ala Val Leu Glu Asn Leu Ala Thr Arg Leu Arg Pro Phe Leu  
 165 170 175  
 35 Gln Cys Tyr Leu Lys Arg Leu Cys Gly Leu Asp Asp Leu Cys Ser Arg  
 180 185 190  
 Arg Arg Leu Ser Asp Ile Lys Asp Ile Ala Ser Phe Val Leu Val Ile  
 195 200 205  
 Leu Ala Arg Leu Ala Asn Arg Val Glu Arg Gly Val Ser Glu Ile Asp  
 40 210 215 220  
 Tyr Thr Thr Val Gly Val Gly Ala Gly Glu Thr Met His Phe Tyr Ile  
 225 230 235 240  
 Pro Gly Ala Cys Met Ala Gly Leu Ile Glu Ile Leu Asp Thr Gln Glu  
 320

245                      250                      255  
 Cys Ser Ser Arg Val Cys Glu Leu Thr Ala Ser His Thr Ile Ala Pro  
                     260                      265                      270  
 Leu Tyr Val His Gly Lys Tyr Phe Tyr Cys Asn Ser Leu Phe  
 5                      275                      280                      285

## (2) INFORMATION FOR SEQ ID NO:102:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 332 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 15 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

Met Leu Ala Val Arg Ser Leu Gln His Leu Thr Thr Val Ile Phe Ile  
 20    1                      5                      10                      15  
 Thr Ala Tyr Gly Leu Val Leu Ala Trp Tyr Ile Val Phe Gly Asp Leu  
                     20                      25                      30  
 His Arg Cys Ile Tyr Ala Val Arg Pro Ala Gly Ala His Asn Asp Thr  
                     35                      40                      45  
 25    Ala Leu Val Trp Met Lys Ile Asn Gln Thr Leu Leu Phe Leu Gly Pro  
                     50                      55                      60  
 Pro Thr Ala Pro Pro Gly Gly Ala Trp Thr Pro His Ala His Val Cys  
 65                      70                      75                      80  
 Tyr Ala Asn Ile Ile Glu Gly Arg Ala Val Ser Leu Pro Ala Ile Pro  
 30                      85                      90                      95  
 Gly Ala Met Ser Arg Arg Val Met Asn Val His Glu Ala Val Asn Cys  
                     100                      105                      110  
 Leu Glu Ala Leu Trp Asp Thr Gln Met Arg Leu Val Val Val Gly Trp  
                     115                      120                      125  
 35    Phe Leu Tyr Leu Ala Phe Val His Gln Arg Arg Cys Met Phe Gly Val  
                     130                      135                      140  
 Val Ser Pro Ala His Ser Met Val Ala Pro Ala Thr Tyr Leu Leu Asn  
 145                      150                      155                      160  
 Tyr Ala Gly Arg Ile Val Ser Ser Val Phe Leu Gln Tyr Pro Tyr Thr  
 40                      165                      170                      175  
 Lys Ile Thr Arg Leu Leu Cys Glu Leu Ser Val Gln Arg Gln Thr Leu  
                     180                      185                      190  
 Val Gln Leu Phe Glu Ala Asp Pro Val Thr Phe Leu Tyr His Arg Pro



195                      200                      205  
 Ala Val Gly Val Ile Val Gly Cys Glu Leu Leu Leu Arg Phe Val Gly  
 210                      215                      220  
 Leu Ile Val Gly Thr Ala Leu Ile Ser Arg Gly Ala Cys Ala Ile Thr  
 5 225                      230                      235                      240  
 Tyr Pro Leu Phe Leu Thr Ile Thr Thr Trp Cys Phe Val Ser Ile Ile  
 245                      250                      255  
 Ala Leu Thr Glu Leu Tyr Phe Ile Leu Arg Arg Asp Ser Ala Pro Lys  
 260                      265                      270  
 10 Asn Ala Glu Pro Ala Ala Pro Arg Gly Arg Ser Lys Gly Trp Ser Gly  
 275                      280                      285  
 Val Cys Gly Arg Cys Cys Ser Ile Ile Leu Ser Gly Ile Ala Val Arg  
 290                      295                      300  
 Leu Cys Tyr Ile Ala Val Val Ala Gly Val Val Leu Met Ala Leu Arg  
 15 305                      310                      315                      320  
 Tyr Glu Gln Glu Ile Gln Arg Arg Leu Phe Asp Leu  
 325                      330

## (2) INFORMATION FOR SEQ ID NO:103:

20

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 482 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

25

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

30

Ala Ala Phe Asp Leu Glu Val Pro Gly His Arg Pro Phe Ala Pro Gly  
 1                      5                      10                      15  
 Pro Ala Leu Pro Pro Gly Gly Leu Ala Val Gly Gly His Met Tyr Val  
 20                      25                      30  
 35 Asn Arg Asn Glu Ile Phe Asn Ala Ala Val Thr Asn Ile Ile Leu Asp  
 35                      40                      45  
 Leu Asp Ile Ala Leu Lys Glu Pro Val Pro Phe Pro Arg Leu His Glu  
 50                      55                      60  
 Ala Leu Gly His Phe Met Arg Gly Ala Ala Val Xaa Leu Leu Phe Pro  
 40 65                      70                      75                      80  
 Ala Ala Arg Val Asn Pro Asp Ala Tyr Pro Cys Tyr Phe Phe Lys Ser  
 85                      90                      95  
 Ala Cys Arg Pro Arg Ala Pro Pro Val Cys Ala Gly Asp Gly Pro Ser

322

	100		105		110
	Ala Gly Gly Asp Asp Gly Asp Gly Asp Trp Phe Pro Asp Ala Gly Gly				
	115		120		125
5	Asp Asp Gly Asp Glu Glu Trp Glu Glu Asp Thr Asp Pro Met Asp Thr				
	130		135		140
	Thr His Gly Pro Leu Pro Asp Asp Glu Ala Ala Tyr Leu Asp Leu Leu				
	145		150		155
	His Glu Gln Ile Pro Ala Ala Thr Pro Ser Glu Pro Asp Ser Val Val				
	165		170		175
10	Cys Ser Cys Ala Asp Lys Ile Gly Leu Arg Val Cys Leu Pro Val Pro				
	180		185		190
	Ala Pro Tyr Val Val His Gly Ser Leu Thr Met Arg Gly Val Ala Arg				
	195		200		205
	Val Ile Gln Gln Ala Val Leu Leu Asp Arg Asn Phe Val Glu Ala Val				
15	210		215		220
	Gly Ser His Val Lys Asn Phe Leu Leu Ile Asp Thr Gly Val Tyr Ala				
	225		230		235
	His Gly His Ser Leu Arg Leu Pro Tyr Phe Ala Lys Ile Gly Pro Asp				
	245		250		255
20	Gly Ser Ala Cys Gly Arg Leu Leu Pro Val Phe Val Ile Pro Pro Ala				
	260		265		270
	Cys Glu Asp Val Pro Ala Phe Val Ala Ala His Ala Asp Pro Arg Arg				
	275		280		285
	Phe His Phe His Ala Pro Pro Met Phe Ser Ala Ala Pro Arg Glu Ile				
25	290		295		300
	Arg Val Leu His Ser Leu Gly Gly Asp Tyr Val Ser Phe Phe Glu Lys				
	305		310		315
	Lys Ala Ser Arg Asn Ala Leu Glu His Phe Gly Arg Arg Glu Thr Leu				
	325		330		335
30	Thr Glu Val Leu Gly Arg Tyr Asp Val Arg Pro Asp Ala Gly Glu Thr				
	340		345		350
	Val Glu Gly Phe Ala Ser Glu Leu Leu Gly Arg Ile Val Ala Cys Ile				
	355		360		365
	Glu Ala His Phe Pro Glu His Ala Arg Glu Tyr Gln Ala Val Ser Val				
35	370		375		380
	Arg Arg Ala Val Ile Lys Asp Asp Trp Val Leu Leu Gln Leu Ile Pro				
	385		390		395
	Gly Arg Gly Ala Leu Asn Gln Ser Leu Ser Cys Leu Arg Phe Lys His				
	405		410		415
40	Gly Arg Ala Ser Arg Ala Thr Ala Arg Thr Phe Leu Ala Leu Ser Val				
	420		425		430
	Gly Thr Asn Arg Leu Cys Ala Ser Leu Cys Gln Gln Cys Phe Ala				
	435		440		445

Thr Lys Cys Asp Asn Asn Arg Leu His Thr Leu Phe Thr Val Asp Ala  
 450 455 460  
 Gly Thr Pro Cys Ser Arg Ser Ala Pro Ser Ser Thr Ser Arg Pro Ser  
 465 470 475 480  
 5 Ser Ser

## (2) INFORMATION FOR SEQ ID NO:104:

- 10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 10212 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

20 GAAAGTGGAG GGAAGGGGGG GAGGAGTGGG AGTGGTTAAG GAGGTGGGAG GAGAGAGAGA 60  
 AGGGAGGGGA GGTGGTGGAG ACTGTGGAGT AAGGAAGGAG AAAGGAGGAG AAGGGAGGGA 120  
 GTTTGATTGT AGAGAAAAG AGGTGGGAGA GGAGAGGGGG ATGATGGATG GAGGGTATGG 180  
 GAGTGGAGGG GGGGGGTGGA GGTGGAGTGG GGGTGGGAGT TAGGGGGTGG GGGGGTGGGG 240  
 GNTGTGGTGG GTTGGTGGTG GGGGGGTGGT GTTGGGTGGG NTGGGGTGGG TGGGTGGGGG 300  
 TTTTTTTTGT TTTTTTTTGT TTTTGTNTNT NNTTTTNNNT NNNNNTNNTN NNNNNTTTTT 360  
 25 TGGCGCCGGA CCTCACGGAC CCGCTGCTGT TTGCGTACGT CGGATTCCAG GTCGTGAACC 420  
 ACGGGCTGAT GTTTGTGGTC CCCGACATCG CCGTATACGC GATGCTGGGG GCGCCCGTGT 480  
 GGATCTCGCT GACGCAGGTG CTTGGGCTCC GCGCCGCCT TCACAAGGAC CCAGACGCCG 540  
 GGCCCTGGGC GGCCGCGACC CTGCGGGGCC TCTTTTCTC CGTCTACGCA TTGGGGTTTG 600  
 CGGCGGGGGT GCTGGTGCGG CCGCGGATGG CCGCGAGCCG GCGGTCGGGG TGATCGCCAT 660  
 30 TTCAAATAAA AGGCACGAGT TCCCCGAATA CCACCGCGCT GTGATGATTT CGCCCTACCG 720  
 CTCCGATCCC CGGGGGGAGG GGGGAAGGAA ATGGGGGCGG GGGTGCCGTG GACGGGTATA 780  
 AAGGCCAGGG GGGCAGGCGG GCCCATCACT GTTAGGGTGT TAGGTTGGGA GGTGGCACAA 840  
 AAAGCGACAC ACCCGTGTG TAGTTGTCCG CGGGAGGCGG TGGTTTCCGG CAACCCTCCT 900  
 CGCTGCGCCG GCGCGCCTCA CCGGTCC'TTC GCGGGGCGG GGGCTCTTCT GGTGATGGCC 960  
 35 CTTGGACGGG TGGGCCTAGC CGTGGGCCTG TGGGGCCTGC TGTGGGTGGG TGTGGTCTGT 1020  
 GTGCTGGCCA ATGCCTCCCC CGGACGCACG ATAACGGTGG GCCC GCGGGG GAAAGAGAGC 1080  
 AATGCCGCCC CCTCCGCGTC CCCGCGGAAC GCATCCGCCC CCCGAACCAC ACCCAGCCCC 1140  
 CCCCACCCC GCAAGGCGAC AAAAAGTAAG GCCTCCACCG CCAAACCGGC CCCGCCCCC 1200  
 AAGACCGGGC CCCCAGAC ATCCTCGGAG CCGGTGCGAT GCAACCGCCA CAACCCGCTG 1260  
 40 GCCCCGTACG GCTTGCGGGT GCAAATCCGA TGCCGGTTTC CCAACTCCAC CCGCACGGAG 1320  
 TCCCGCCTCC AGATCTGGCG TTATGCCACG GCGACGGACG CCGAGATCGG AACGGCGCCT 1380  
 AGCTTAGAGG AGGTGATGGT AAACGTGTCG GCCCCGCCC GGGGCCAACT GGTGTATGAC 1440  
 AGCGCCCCCA ACCGAACGGA CCCGCACGTG ATCTGGGCGG AGGGCGCCG CCCGGGCGCC 1500

	AGCCCGCGGC	TGTACTCGGT	CGTCGGGCGG	CTGGGTCGGC	AGCGGCTCAT	CATCGAAGAG	1560
	CTGACCCTGG	AGACCCAGGG	CATGTACTAC	TGGGTGTGGG	GCCGGACGGA	CCGCCCCTCC	1620
	GCGTACGGGA	CCTGGGTGCG	CGTTCGCGTG	TTCCGCCCCC	CGTCGCTGAC	CATCCACCCC	1680
	CACGCGGTGC	TGGAGGGCCA	GCCGTTTAAG	GCGACGTGCA	CGGCCGCCAC	CTACTACCCG	1740
5	GGCAACCGCG	CGGAGTTCGT	CTGGTTCGAG	GACGGTCGCC	GGGTATTCTGA	TCCGGCCCAG	1800
	ATACACACGC	AGACGCAGGA	GAACCCCGAC	GGCTTTTCCA	CCGTCTCCAC	CGTGACCTCC	1860
	GCGGCCGTCG	GCGGCCAGGG	CCCCCGCGC	ACCTTCACCT	GCCAGCTGAC	GTGGCACCCG	1920
	GACTCCGTGT	CGTTCTCTCG	GCGCAACGCC	AGCGGCACGG	CATCGGTGCT	GCCGCGGCCA	1980
	ACCATTACCA	TGGAGTTTAC	GGGCGACCAT	GCGGTCTGCA	CGGCCGGCTG	TGTGCCCAG	2040
10	GGGGTGACGT	TTGCCTGGTT	CCTGGGGGAC	GACTCCTCGC	CGGCGGAGAA	GGTGCCCGTC	2100
	GCGTCCCAGA	CATCGTGC GG	GCGCCCCGGC	ACCGCCACGA	TCCGCTCCAC	CTTGCCGGTC	2160
	TCGTACGAGC	AGACCGAGTA	CATCTGCCGG	CTGGCGGGAT	ACCGGACGG	AATTCCGGTC	2220
	CTAGAGCACC	ACGGCAGCCA	CCAGCCCCCG	CCGCGGGACC	CCACCGAGCG	GCAGGTGATC	2280
	CGGGCGGTGG	AGGGGCGGG	GATCGGAGTG	GCTGTCTCTG	TCGCGGTGGT	TCTGGCCGGG	2340
15	ACCGCGGTAG	TGTACCTCAC	CCACGCCTCC	TCGGTGC GCT	ATCGTCGGCT	GCGGTAATC	2400
	CGGGGCCGGG	CCCGGCCGGC	GGTTGTCTTC	TTTTCCACCC	CTTCCGTCCC	CCGTACCCAC	2460
	CACACCCAC	CCCACCCCC	CGCCGTCCCC	CGGGCGTTAT	AAGCCGCCGC	ACTCGCTTTT	2520
	CCCACCGGAA	AATCCTCGGC	CCGATCCGAA	CGGCGCACGC	CGCGTGGGCT	CCAAACGCCT	2580
	CCGCGAAGAG	AGCGCCCCGC	CCCGATATTC	AAGCCCGCGG	TGGTGCTATG	GCTTTCCGTG	2640
20	CTTCGGGACC	CGCCTACCAG	CCCTCGCCCC	CCGCGGCCCT	CCCGGCGCGG	GCTCGTGTTT	2700
	CGGCCGTGGC	CTGGATCGGC	GTCGGAGCGA	TCGTGCGGGC	CTTTGCGCTC	GTCGCCCGCT	2760
	TGGTTCTCGT	ACCCCTCGG	TCCTCGTGGG	GACTCTCGCC	GTGCGACAGC	GGCTGGCAGG	2820
	AATTCAACGC	GGGATGCGTC	GCGTGGGACC	CCACCCCGCT	CGAGCACGAG	CAGGCGGTCTG	2880
	GCGGCTGCAG	CGCGCCGGCC	ACCCTTATCC	CCCGTGC GGC	CGCCAAGCAC	CTGGCCGCTC	2940
25	TGACACGCGT	CCAGCGGAG	AGATCGTCGG	GTTACTGGTG	GGTGAACGGA	GACGGCATCC	3000
	GGACCTGTCT	GAGACTCGTC	GACAGCGTCA	GTGGCATCGA	CGAGTTTTCG	GAGGAGCTCG	3060
	CGATCCGCAT	ATGCTACTAC	CCACGAAGCC	CCGGCGGGTT	TGTCCGCTTC	GTAACCTCAA	3120
	TACGTAACGC	CCTGGGGTTG	CCGTGAGGCG	CGCGTCCGAC	GGTCCCCTTT	CTCGCCTCTC	3180
	TTCTTCCCCC	TCCCCACCCC	ACCCACCGAC	CAACGACGGC	GTTTGGCCAA	TACCCTCCTT	3240
30	TTTTCTTTT	CTCTTCCCCC	CCCAAAAAA	AAAACAATAA	ACAGCTAATT	GCGTACGACA	3300
	AACCATGCGG	AACCTCGCTG	TTTTTTTCCC	CTGTTTGTTA	CTTTTATTG	AAAACAGACA	3360
	TACGGGGAAA	GGGGCCGGA	ACCGAGACGG	TGGGGCCGGC	GGTCGCATTT	TTTAAATGGC	3420
	TCTGGTGTCTG	GCCGCGTTTG	AGCTTCGTCA	ACAGGGCGCT	GAGGGCGGCG	ACGTTTGTCTG	3480
	GGCCGTCTGTT	GGCCAGCGCG	TTGGTCCGGG	GGCGGGCGGG	CATGGGCGAC	AGGCTTAGTC	3540
35	CCGGGTCCGG	GGCGCGTGTG	GCCCCCGGAG	GGGAGAAGAG	GGCAGACCCG	CCCCAGTCGT	3600
	ACAGGGGATT	TTCCGCCTCG	ATGTACGGGG	AGTCCGGGGC	GTCTCCCGGC	GGGGCCGCC	3660
	CGCCGGCGTC	TTGCCGGCGA	AGGCAGATGT	TTTCGTATAC	CCGAACCCAG	GGGATCTCCT	3720
	CGTAGACGCG	CCCCCATCC	TCGCTCACCG	ACTCGTAAAT	GGAATCTGCC	TCCTCGGAGG	3780
	GGGCGCGGGG	GGCGTGGCTT	TCGGCCGGCC	AGGCGGCGGC	GGTGGTGTCTG	GCGGCGGGGG	3840
40	TGGAGCCAAG	CCCGACGCC	GCGGGCATGG	CGGCGTCATC	CTCCGGCAGC	AGATACGTGT	3900
	TTTCCATCTG	GTCCGGTTCG	GCCTCCGCGT	CTGGCCCCCA	GGTCCGCACT	GCGTTGTAAA	3960
	CCCCGGCGGC	CTCGCGATGA	GCCGCGAGCG	GGCGCGCCGC	GGCTGCCGGC	CGCTGCTCGG	4020
	GGGGCGCGGG	GTTGCGGGGC	GGGAGGCGCG	GGGGCGCCCC	GGCCATATGC	GTGTAATAAG	4080

	TGGCCGCGCG	GCCGGCGCAG	GGCTCGGGAC	CCCGTCCGGC	CGCGTCAACG	TGCGGGGGCT	4140
	CGGGGAGGTC	CTCGCGGTGG	CGCCTGAACC	TCCGAGGGGC	CGCGGGGGTC	AATTGGGGGC	4200
	CACCCCGGGG	GAGCGGCGGG	GGTGCCTTAT	CGCGCCGGGT	CCGTGTATATC	TTGTCCCGGC	4260
	AGCTCCCGCC	GACCGCGCCG	CGGCCCCCGG	GTGGGCCGGA	CGCCGCGAGG	CGCAGGATGG	4320
5	ACTCGTAGTG	GGGCGACGGG	GTTCCGCTCC	GAAGCAGGTC	CGGGGCCAGG	CGCGCCCCCTA	4380
	ACCACTACTT	GATGCTGAGT	TCCATCCGGG	CCCAGCTCGG	GGCGGTTCATC	GTGGGGAACA	4440
	GGGGGCGGCG	GGTCCTGCAG	AAGCGCTCCT	GGCTGTCCAC	CGCCGCCCGT	AGGTACTCGT	4500
	TGTTTAGGCT	GTCGGAGGCC	CAAACAACAT	ACCCGGTAAG	CGTCGCGTTA	ATTATATACT	4560
	GGGCGTGGTG	GTGGAATATG	GATAGAACCT	CGACGGTCGA	GACGATGGCG	TCCACGATCC	4620
10	CGTACGTGCC	GCCGCTGCGC	TTGCCGGTCT	CCCACAGGTG	GGCCAGGCGC	GTCAGGTGGC	4680
	CCAGGACGTC	GCTGACCGCC	GCCCGCAGGG	CCATGCACTG	CATCGAGCCC	GTGGTGCCGC	4740
	TGGGCCCGCG	GTCCAGGTGG	CGCGCAAACG	TCTCCGCGGG	CGCCTCCAGA	CTCCCGCTGA	4800
	GCGCCACGAA	CCGCGGATCG	GCGGGGCCCCA	GGCGGCGACA	CACGTACTTG	TCCGCCGTCC	4860
	ACAGCATCCA	CGAGGCCCAA	TGGTACAACA	CGGAGACGTA	GGCCAGGAGC	TCGCTCAGCC	4920
15	GCA GTGCGGT	GTCCGTGCTC	GGCCGGCTCG	GGTCTGCGGG	GCGCATAAAA	AACATGTACT	4980
	GCTGGAGCCT	GTGGGCCGCG	TCGCGCAACC	CCGCCACCGC	GGCGGCGTAC	TTGGCCGCGG	5040
	CGGCCCCGCT	CTTGAACGGG	GCGCGCACCA	CCAGCTTCGG	GAGCAGGGTG	GGCCGCATCA	5100
	ACACGTGCAG	GCTGGGGTCG	CANTCGCCCG	CCGGGTCTGC	GGGGATGTCC	AGGCCGCTGG	5160
	GCACAACCGT	CTGGAGGTAC	TTCCAGTACT	GCGCTAGGAT	GGCGCGGCTC	AGCTGGCCGC	5220
20	CCGACAGCTC	CACCTCGCCG	AGCGCCTGCT	TGGCGGCCGA	CGCGTAGTGC	CGGATGTAGT	5280
	CGTAGTGCGG	GTCGCTGGCG	AGCCCGTCTA	CGATCAGGCT	CTCGGGGACG	GTGTTATGGT	5340
	GCCGCGCCGC	CAGCCGGACG	CTGCGATCGG	CGCCGGTCAG	AAACGCCCGC	TGCAGGTCGT	5400
	CGGCGCGCTG	CCGACGGACG	CCCACGGCCG	CGCTGAGGAG	CCCCTCCGGG	GTGGGGAGCA	5460
	GACACCCGGC	GAAGATGCGC	CGCTCGGGGA	CGCCCGCGTT	GGCGCCGCGG	ATGAGGTTGG	5520
25	CCGGCGTCAG	GCACCGCGCC	AGCCGCAGGG	AGCTCGCGCC	GCGCGCCCGG	CGTTGCATGG	5580
	CGGAGACCGT	TCGGTCCGGG	GCCCGCCGGT	CGGAGGTATG	CCGCGTCCCG	GGATATAGGG	5640
	TTGCTTTTTA	TGGGGAGGCG	CCTATGGGCG	TGGCGGGCCG	CCCAGCCCGG	TCGCGCGCCT	5700
	CCCGGACACG	TGCGCCCGGA	GGGCGGCGGT	CTCCTCGTCC	CCCATGAGCA	GTTTCCGAAA	5760
	CTGCGCCATG	ATGTCCACGA	CGCGGACCCG	CGGCCCCAGC	ACGGACTCGC	TATTCAGGGG	5820
30	GGCGGGGGGG	AAGGCCGCCA	GGTCTTCGAG	CAGGAAGGCG	GGGTCTGCCG	TCCCGCTCAC	5880
	GGGCGCCCGG	GGCGCCGAGG	ACGCGGGGCG	AAGGTCCACG	TGTTCCGCGG	CGGCGCGCAC	5940
	GTCCGCCCAA	AATTTGGCGG	GGGTGGTCCG	CGCGTACAGG	GGCTGGGTCC	CGCGGAGGAC	6000
	GCACGCGTAG	CGCAGGGGGG	TGTACGTGCC	CACCTCGGGG	GCCGTCCGACC	CGCCGTCAAA	6060
	CGCGGCCAGG	GCCACGCACG	CGACCACCGT	GTCGGCCAGG	CCCAGCAGCC	GCTGCAGGAT	6120
35	GAGCCCCGTC	GCCAGCACGG	CGCGCGCGGC	CGCCGCGTTG	TCCCTGCGCC	GGCGCGCGTC	6180
	CCCGCAGGCC	AGGGCGTATT	TCAGGGTAAC	GGTCGCCAGG	GCCGTGTGCA	GCGCGTACAC	6240
	GGCCGCGCCC	AGCACGGCGT	TCAGCCCGCT	GGTGGCGAGC	AGGCGGCGCG	CCGCGGTGTC	6300
	GCCCAGCGCC	TCGTGCTCGG	CCGCCACGAC	CCCGGGGCTA	CCCAGGGGCA	GGGCGCGAAA	6360
	CAGCGCCTCC	TGCTCCACGT	CCGCAAACGC	GGGGTGGGCG	GAGTGCGGGT	GCAGGCGCGC	6420
40	CCCCACGACC	ACCGAGAGCC	ACTGGACCGT	CTGCTCCGCC	AGGACCGCCA	GCACGTCCAG	6480
	GACGCGCCCC	GCAAACGCGG	CCTCCCGCGG	GAGCACGCAT	TTGACGGCGC	CGGGGTGAA	6540
	GCGGGCGAGC	AGAGCCCCCG	TGGCGATGTA	CGTCATGCGC	CCCGCGTAGC	GGGCGGCCAC	6600
	GCGACAGTCG	CGCCCCAGGA	GCGCGCGCAC	CCCGGGCCAG	TACAGCAGGG	ACCCCAGCGA	6660

	ACTGCGAAAG	ACCGCGGCGT	CGGGGCCGGG	GTGGGGGGGC	GCGGCCCTC	CCGCGCTGAG	6720
	CAGCGGCACG	GCGGCGGCC	CCACGGGCCG	CAACGCCGTG	AGGCTCGCGA	ACTGCCGTG	6780
	GAGCTCGGCC	GCCCTGTCTG	CGAGCTCCGA	GCCGCGCCCC	TCCGTGTGCA	GGCGCGTCCC	6840
	GCAGACCCAC	CCGTTGATCG	CCACCCGCAC	GATGGCGTCC	ACCAGAAAGC	CCATCGCGCG	6900
5	GGAGGGGCTG	GTTTTTGCCC	GCCGATCCGT	CAGGTCGAGG	ATCGCGTTCG	CCGTGACGTA	6960
	CCAGGCCAGC	GCCTCGCCCT	GCTGCAGCGT	CTGGCGGAAA	AACACCTTTG	GGTCGGCCGG	7020
	GGAGGCAAAG	TGCATGACCC	CCACGCGCGA	CAGCCCGAAC	GCGCTATCCG	GACACGGGTA	7080
	GAACCCGGCC	GGATGTCCCA	GGGCCAGGGC	CGAGCGCACG	GACTCGTCCC	ACGCGGCGAC	7140
	TCGGGGGGTC	AGGCGGTCCA	GGGGGAATGC	CGCCTGCAGC	TCCGGGCCCG	ACACGCGGCC	7200
10	CTCTATAATC	TCGACCGTCG	CGGGAGGCCG	CGCCCCGGCG	CCGTCATCGT	GCGCGACGGC	7260
	GGCGGGGTAG	TCGTCTCCT	CGTAGCTGAG	CTCGTCCAGG	AACAGCGGCG	AGGGCACCAC	7320
	CCGCGAACCG	CCACCCGCC	CCAAAACGTC	GCGTGGGTCC	ATCGGGCCCA	GGTAGCCTCC	7380
	CCGCGGGGCC	CGCGTGATGG	CGCTGTCCCG	GCGTCCGCGA	ACGGA CTGGC	TCCTGGCCGT	7440
	AACGGACCTG	GGGCGCGGAA	AGGACGCCCC	GCGGGGGGGC	GCCGCCGCC	GGGCCTCGGA	7500
15	CGCGCGTCGG	GACCCGGGGT	GACCGCGGGC	CTCCCGGCGA	CGGCGCGGGG	GCGGCTCTTC	7560
	GCTCGCCATC	TCCCCCGCGG	CCTCGACCTC	GCTGTCTGTC	TCCACGTTAA	ACACCGCCCC	7620
	CAGGTACCCC	ATTAACCCGA	CTCCACCGCC	CTCGGGCTCG	TCCTCCACGG	GCGATTGCGC	7680
	GCGATGCGCG	GACGGGGCAT	GGGACCGGGT	GGAGGCGCGC	CTCCGGCGTA	CGGCATGCCC	7740
	GCGCACGGAC	ATGGTGGCCG	GAGGCCCCAT	TTTTTACACA	CCCCCTCCCC	GCAAACGGAC	7800
20	AAGGAAAGGG	GTGGTGGGAG	GGGGGAGGCC	CAAACGGGGA	GGTGGGGGGT	AGGGGGCGGT	7860
	CCCAGGGAGC	GGGGGTACG	AACCGGCACG	ACGGGAACAG	AGAAACGCGA	CCGCTCCAAC	7920
	AAGGGTGGGG	GGTGGGCCCT	ATCCCCACGC	AAACCCGCGG	GCAAATGCGA	GAACGGGACC	7980
	CGCGCGCCTG	CCTTTATACG	CGGACCCAG	CACCACGAGC	CGTTCTGTGA	CCCGAATCTA	8040
	CACGACCGCG	GGCTCGTAGG	CGCGACTAAC	GCCCAACCCA	ACGGCACACA	CCCCCACC	8100
25	CGCGCGTAAC	CCCATTTCTT	TCATGGTCCC	GTAATAAACA	GCCAACGCAC	GCCGCGTATG	8160
	ATGAGTTGCT	TGCCAATGTT	TATTGCTGTG	GTTGCGAACC	CTCTATCGCG	ATACAGACGG	8220
	AGGTGAGGCG	GGGCGGTGGT	GGGGGGGGGG	GCGCCCCCCC	CGGTGCGACA	TCCTACCCCC	8280
	CAAAGTCGTC	AATGCCCATG	GCATCGGTAA	ACATCTGTTC	AAACTCAAA	TCGTCCACGT	8340
	CCAAAGCCCC	ATACAAAACG	GGGTCGTGGG	TCATTCCCGG	GGAGGGGGAC	TCCACGTCCC	8400
30	CCAGCATCTC	CAAGTCGAAG	TCGTCCAGGG	CGTCGGCGGG	CGTCATATCC	ACCTCTCTCG	8460
	CGTCCAGGCG	GAGTTCGTCT	CCCAGGCTGA	CGTCGGTAAT	GGGGGCGGTG	GTGGACAGTC	8520
	TGCGGGGGCG	TTGTCCCGCG	GAGAGAAACG	ACATGCGCGG	CGCCACCAGC	CCGGCCTCCG	8580
	CGGGAGCGTC	ATCGTCGTCC	GGGAGGTCGA	GCAGGCCCTC	GATTGTGCGAT	CCGTAATTAT	8640
	TTCTGGTCCG	CCCGCGGCTA	TACGCGTGCT	CCCGCATGAC	GGA CTGCCC	TCCGAGGTG	8700
35	CAACGCTGGA	GTACGAGTCC	AAC TTGGCCC	GGATCAGCAG	CATAAAGTAC	CCAGAGGAGC	8760
	GGGCCTGGTT	GCCCTGCAGG	ACGGGCGGGG	TCGTGAGGGG	CGCCCCGGGT	TCCTCCGCCG	8820
	CCGCACTTCG	CACCAGCGGG	AGGTTCAAGT	GCTCGCGAAT	GTGGTTTAGC	TCCCGCAGTC	8880
	GCCGGGCCCTC	CACGGGAACT	CCCCGCACGG	TGAGCGATCC	GTTGATAAAC	ATCAGGGGCT	8940
	GAAACAGACA	CGCCAACTGG	CGCCAGCTCT	CCAGGTGCGA	GCAGAGGCCG	TCGAACAGAT	9000
40	CGGGCCGCAT	CATCTGCTCG	GCGTACGCGG	CCCATAGGAT	CTCGCGGCTC	AGAAAGAGGT	9060
	ATAGATGCAG	AAACAGGACG	CGCGCCAGGC	GCGCGGTCTC	GCGGTAGTAC	CTGTCCGCGA	9120
	TCGTGGTGCG	CAGCATCTCC	CGCAGGTTCG	GGTGCGGGCC	CCGCATGTGT	GCCTGGCGGT	9180
	GTAGCTGCCG	AACGCTGGCG	CGCAGGTACC	GGTACAGGGC	CGAGCAAAAA	TTTGCCAACA	9240

CGGTCCGGTA GCTCTCCTCC CGCGCCCGCA GCTCACCGCG GAAAAACTGC GCCATGGCCT 9300  
 CGTAGTACGA AGGCAGCTCG TCGCGGGTGG CGGGCAGGGT GGGGAACGCC ACGTCGCCGT 9360  
 GGGCGCGAAT GTCGATCGGG GAGCGCTCGG GGACGTGCGC ATCCCCCAG TCGATCACGT 9420  
 CGCTGGGCAG CGTCGACAGA AACTTGCACT CCCGGTACAT GTCGGCGTTG GTCGGGAACC 9480  
 5 CAGAGAACAG GTCCTCGTTC CAGGTATCTA GCATGGTACA CAGCGCGGGA CCCGCGCTGA 9540  
 AGCCCAGATC GTCGAGGAGA CGGTAAACA GGGCCGCGGG GGGGACGGGC ATGGGCGGCG 9600  
 AGGGCATCAG CTGGGCCTGA CTCAGCCGAC CGGTGGCGTA CAGCGGAGGG GCGGCTGGGG 9660  
 TGTTCCTGGG ACCCCCGGCT GGCCTGGGGG GCGGTGGCGA AACCCCGTCC GCGTCCGCAA 9720  
 ACAGATTGTT GACCAACAGG TCCATGGGGG CGGTTGGGTC CGGGGATAAC GATTTTGAGA 9780  
 10 GGCGAATGAG AAGTGCCCGA GCGCCCGCG GCGGAGAGGG GGGGAGGGAT CCGGGACCCG 9840  
 CGACAGAAAA AGGCCGGGGC CCTTGCGAAG GGAATTGCCG GGGGTGCCGT GCGTCCCCGA 9900  
 TGA CTGACAT CTCTCTTCTT CCCCCCGCA TTTT TAGTAT CACCCCAATT GCCGCCCAA 9960  
 AACCTTCTTG ACTTCCCCCA CCCGTTTCCG TGGCGGCCCC TTCCCCCTG CTCCTCTGTA 10020  
 ACGGGATGGT CTTATTCCCT CCTTCCCCTG GCCCTTCCCC CTCCTCTCTT CCTTTTCTCT 10080  
 15 TCCCTTCTT CCGTCACTCC TTCCTCCCCT CTCTCGATTC CTCCCTTCTT CCCCATCTCT 10140  
 TCCTTCTCT CTCACTCTCA TATCCTTCAA TACTCTCCTC CTCTCTATCT TTCCCCCGC 10200  
 TTCTTCTCTC T 10212

## (2) INFORMATION FOR SEQ ID NO:105:

20

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 148 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

25

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

30

Val Gly Val Gly Val Arg Gly Trp Gly Gly Gly Xaa Cys Gly Gly Leu  
 1 5 10 15  
 Val Val Gly Gly Trp Cys Trp Val Xaa Trp Gly Gly Trp Val Gly Val  
 20 25 30  
 35 Phe Phe Cys Phe Phe Leu Phe Cys Xaa Xaa Phe Xaa Xaa Xaa Xaa  
 35 40 45  
 Xaa Xaa Phe Leu Ala Pro Asp Leu Thr Asp Pro Leu Leu Phe Ala Tyr  
 50 55 60  
 Val Gly Phe Gln Val Val Asn His Gly Leu Met Phe Val Val Pro Asp  
 40 65 70 75 80  
 Ile Ala Val Tyr Ala Met Leu Gly Gly Ala Val Trp Ile Ser Leu Thr  
 85 90 95  
 Gln Val Leu Gly Leu Arg Arg Arg Leu His Lys Asp Pro Asp Ala Gly

100 105 110  
 Pro Trp Ala Ala Ala Thr Leu Arg Gly Leu Phe Phe Ser Val Tyr Ala  
 115 120 125  
 Leu Gly Phe Ala Ala Gly Val Leu Val Arg Pro Arg Met Ala Ala Ser  
 5 130 135 140  
 Arg Arg Ser Gly  
 145

## (2) INFORMATION FOR SEQ ID NO:106:

10

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 538 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 15 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

20

Met Gly Ala Gly Val Pro Trp Thr Gly Ile Lys Arg Ala Gly Gly Pro  
 1 5 10 15  
 Ile Thr Val Arg Val Leu Gly Trp Glu Val Ala Gln Lys Ala Thr His  
 20 25 30  
 25 Pro Cys Cys Ser Cys Pro Arg Glu Ala Val Val Ser Gly Asn Pro Pro  
 35 40 45  
 Arg Cys Ala Gly Arg Ala His Arg Ser Phe Ala Gly Ala Gly Ala Leu  
 50 55 60  
 Leu Val Met Ala Leu Gly Arg Val Gly Leu Ala Val Gly Leu Trp Gly  
 30 65 70 75 80  
 Leu Leu Trp Val Gly Val Val Val Val Leu Ala Asn Asp Gly Arg Thr  
 85 90 95  
 Ile Thr Val Gly Pro Arg Gly Lys Glu Ser Asn Ala Ala Pro Ser Asp  
 100 105 110  
 35 Arg Asn Ala Ser Ala Pro Arg Thr Thr Pro Thr Pro Pro Gln Pro Arg  
 115 120 125  
 Lys Ala Thr Lys Ser Lys Ala Ser Thr Ala Lys Pro Ala Pro Pro Pro  
 130 135 140  
 Lys Thr Gly Pro Pro Lys Thr Ser Ser Glu Pro Val Arg Cys Asn Arg  
 40 145 150 155 160  
 His Asn Pro Leu Ala Arg Tyr Gly Leu Arg Val Gln Ile Arg Cys Arg  
 165 170 175  
 Phe Pro Asn Ser Thr Arg Thr Glu Ser Arg Leu Gln Ile Trp Arg Tyr



		180		185		190	
	Ala Thr	Ala Thr Asp Ala Glu Ile Gly Thr Ala Pro Ser Leu Glu Glu					
		195		200		205	
	Val Met Val Asn Val Ser Ala Pro Pro Gly Gly Gln Leu Val Tyr Asp						
5		210		215		220	
	Ser Ala Pro Asn Arg Thr Asp Pro His Val Ile Trp Ala Glu Gly Ala						
	225		230		235		240
	Gly Pro Gly Asp Arg Lys Val Val Gly Pro Leu Gly Arg Gln Arg Leu						
		245		250		255	
10	Ile Ile Glu Glu Leu Thr Leu Glu Thr Gln Gly Met Tyr Tyr Trp Val						
		260		265		270	
	Trp Gly Arg Thr Asp Arg Pro Ser Ala Tyr Gly Thr Trp Val Arg Val						
		275		280		285	
	Arg Val Phe Arg Pro Pro Ser Leu Thr Ile His Pro His Ala Val Leu						
15		290		295		300	
	Glu Gly Gln Pro Phe Lys Ala Thr Cys Thr Ala Ala Thr Tyr Tyr Pro						
	305		310		315		320
	Gly Asn Arg Ala Glu Phe Val Trp Phe Glu Asp Gly Arg Arg Val Phe						
		325		330		335	
20	Asp Pro Ala Gln Ile His Thr Gln Thr Gln Glu Asn Pro Asp Gly Phe						
		340		345		350	
	Ser Thr Val Ser Thr Val Thr Ser Ala Ala Val Gly Gly Gln Gly Pro						
		355		360		365	
	Pro Arg Thr Phe Thr Cys Gln Leu Thr Trp His Arg Asp Ser Val Ser						
25		370		375		380	
	Phe Ser Arg Arg Asn Ala Ser Gly Thr Ala Ser Val Leu Pro Arg Pro						
	385		390		395		400
	Thr Ile Thr Met Glu Phe Thr Gly Asp His Ala Val Cys Thr Ala Gly						
		405		410		415	
30	Cys Val Pro Glu Gly Val Thr Phe Ala Trp Phe Leu Gly Asp Asp Ser						
		420		425		430	
	Ser Pro Ala Glu Lys Val Ala Val Ala Ser Gln Thr Ser Cys Gly Arg						
		435		440		445	
	Pro Gly Thr Ala Thr Ile Arg Ser Thr Leu Pro Val Ser Tyr Glu Gln						
35		450		455		460	
	Thr Glu Tyr Ile Cys Arg Leu Ala Gly Tyr Pro Asp Gly Ile Pro Val						
	465		470		475		480
	Leu Glu His His Gly Ser His Gln Pro Pro Pro Arg Asp Pro Thr Glu						
		485		490		495	
40	Arg Gln Val Ile Arg Ala Val Glu Gly Ala Gly Ile Gly Val Ala Val						
		500		505		510	
	Leu Val Ala Val Val Leu Ala Gly Thr Ala Val Val Tyr Leu Thr His						
		515		520		525	

Ala Ser Ser Val Arg Tyr Arg Arg Leu Arg  
530 535

## (2) INFORMATION FOR SEQ ID NO:107:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 170 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

15

Met Ala Phe Arg Ala Ser Gly Pro Ala Tyr Gln Pro Leu Ala Pro Ala  
1 5 10 15  
Asp Ala Arg Ala Arg Val Pro Ala Val Ala Trp Ile Gly Val Gly Ala  
20 25 30  
Ile Val Gly Ala Phe Ala Leu Val Ala Ala Leu Val Leu Val Pro Pro  
35 40 45  
Arg Ser Ser Trp Gly Leu Ser Pro Cys Asp Ser Gly Trp Gln Glu Phe  
50 55 60  
Asn Ala Gly Cys Val Ala Trp Asp Pro Thr Pro Val Glu His Glu Gln  
25 65 70 75 80  
Ala Val Gly Gly Cys Ser Ala Pro Ala Thr Leu Ile Pro Arg Ala Ala  
85 90 95  
Ala Lys His Leu Ala Ala Leu Thr Arg Val Gln Ala Glu Arg Ser Ser  
100 105 110  
30 Gly Tyr Trp Trp Val Asn Gly Asp Gly Ile Arg Thr Cys Leu Arg Leu  
115 120 125  
Val Asp Ser Val Ser Gly Ile Asp Glu Phe Cys Glu Glu Leu Ala Ile  
130 135 140  
Arg Ile Cys Tyr Tyr Pro Arg Ser Pro Gly Gly Phe Val Arg Phe Val  
35 145 150 155 160  
Thr Ser Ile Arg Asn Ala Leu Gly Leu Pro  
165 170

## (2) INFORMATION FOR SEQ ID NO:108:

40

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 215 amino acids  
(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

```

Met Ala Gly Ala Pro Pro Arg Leu Pro Pro Arg Asn Pro Ala Pro Pro
  1           5           10           15
10  Glu Gln Arg Pro Ala Ala Ala Arg Pro Leu Ala Ala His Arg Glu
      20           25           30
    Ala Ala Gly Val Tyr Asn Ala Val Arg Thr Trp Gly Pro Asp Ala Glu
      35           40           45
    Ala Glu Pro Asp Gln Met Glu Asn Thr Tyr Leu Leu Pro Glu Asp Asp
15      50           55           60
    Ala Ala Met Pro Ala Gly Val Gly Leu Gly Ser Thr Pro Ala Ala Asp
    65           70           75           80
    Thr Thr Ala Ala Ala Trp Pro Ala Glu Ser His Ala Pro Arg Ala Pro
      85           90           95
20  Ser Glu Glu Ala Asp Ser Ile Tyr Glu Ser Val Ser Glu Asp Gly Gly
      100          105          110
    Arg Val Tyr Glu Glu Ile Pro Trp Val Arg Val Tyr Glu Asn Ile Cys
      115          120          125
    Leu Arg Arg Gln Asp Ala Gly Gly Ala Ala Pro Pro Gly Asp Ala Pro
25      130          135          140
    Asp Ser Pro Tyr Ile Glu Ala Glu Asn Pro Leu Tyr Asp Trp Gly Gly
    145          150          155          160
    Ser Ala Leu Phe Ser Pro Pro Gly Ala Thr Arg Ala Pro Asp Pro Gly
      165          170          175
30  Leu Ser Leu Ser Pro Met Pro Ala Arg Pro Arg Thr Asn Ala Asn Asp
      180          185          190
    Gly Pro Thr Asn Val Ala Ala Leu Ser Ala Leu Leu Thr Lys Leu Lys
      195          200          205
    Arg Gly Arg His Gln Ser His
35      210          215

```

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:

40

(A) LENGTH: 393 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

5 Met Gln Arg Arg Arg Ala Ser Ser Leu Arg Leu Ala Arg Cys Leu Thr  
1 5 10 15  
Pro Ala Asn Leu Ile Arg Gly Ala Asn Ala Gly Val Pro Glu Arg Arg  
20 25 30  
10 Ile Phe Ala Gly Cys Leu Leu Pro Thr Pro Glu Gly Leu Leu Ser Ala  
35 40 45  
Ala Val Gly Val Leu Arg Gln Arg Ala Asp Asp Leu Gln Pro Ala Phe  
50 55 60  
Leu Thr Gly Ala Asp Arg Ser Val Arg Leu Ala Ala Arg His His Asn  
15 65 70 75 80  
Thr Val Pro Glu Ser Leu Ile Val Asp Gly Leu Ala Ser Asp Pro His  
85 90 95  
Tyr Asp Tyr Ile Arg His Tyr Ala Ser Ala Ala Lys Gln Ala Leu Gly  
100 105 110  
20 Glu Val Glu Leu Ser Gly Gly Gln Leu Ser Arg Ala Ile Leu Ala Gln  
115 120 125  
Tyr Trp Lys Tyr Leu Gln Thr Val Val Pro Ser Gly Leu Asp Ile Pro  
130 135 140  
Asp Asp Pro Ala Gly Xaa Cys Asp Pro Ser Leu His Val Leu Met Arg  
25 145 150 155 160  
Pro Thr Leu Leu Pro Lys Leu Val Val Arg Ala Pro Phe Lys Ser Gly  
165 170 175  
Ala Ala Ala Ala Lys Tyr Ala Ala Ala Val Ala Gly Leu Arg Asp Ala  
180 185 190  
30 Ala His Arg Leu Gln Gln Tyr Met Phe Phe Met Arg Pro Ala Asp Pro  
195 200 205  
Ser Arg Pro Ser Thr Asp Thr Ala Leu Arg Leu Ser Glu Leu Leu Ala  
210 215 220  
Tyr Val Ser Val Leu Tyr His Trp Ala Ser Trp Met Leu Trp Thr Ala  
35 225 230 235 240  
Asp Lys Tyr Val Cys Arg Arg Leu Gly Pro Ala Asp Arg Arg Phe Val  
245 250 255  
Ser Gly Ser Leu Glu Ala Pro Ala Glu Thr Phe Ala Arg His Leu Asp  
260 265 270  
40 Arg Gly Pro Ser Gly Thr Thr Gly Ser Met Gln Cys Met Ala Leu Arg  
275 280 285  
Ala Ala Val Ser Asp Val Leu Gly His Leu Thr Arg Leu Ala His Leu  
290 295 300

Trp Glu Thr Gly Lys Arg Ser Gly Gly Thr Tyr Gly Ile Val Asp Ala  
 305 310 315 320  
 Ile Val Ser Thr Val Glu Val Leu Ser Ile Val His His His Ala Gln  
 325 330 335  
 5 Tyr Ile Ile Asn Ala Thr Leu Thr Gly Tyr Val Val Trp Ala Ser Asp  
 340 345 350  
 Ser Leu Asn Asn Glu Tyr Leu Arg Ala Ala Val Asp Ser Gln Glu Arg  
 355 360 365  
 Phe Cys Arg Thr Ala Ala Pro Leu Phe Pro Thr Met Thr Ala Pro Ser  
 10 370 375 380  
 Trp Ala Arg Met Glu Leu Ser Ile Lys  
 385 390

## (2) INFORMATION FOR SEQ ID NO:110:

15

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 680 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

20

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

25

Met Ser Val Arg Gly His Ala Val Arg Arg Arg Arg Ala Ser Thr Arg  
 1 5 10 15  
 Ser His Ala Pro Ser Ala His Arg Ala Glu Ser Pro Val Glu Asp Glu  
 20 25 30  
 30 Pro Glu Gly Gly Gly Val Gly Leu Met Gly Tyr Leu Arg Ala Val Phe  
 35 40 45  
 Asn Val Asp Asp Asp Ser Glu Val Glu Ala Ala Gly Glu Met Ala Ser  
 50 55 60  
 Glu Glu Pro Pro Pro Arg Arg Arg Arg Glu Arg His Pro Gly Ser Arg  
 35 65 70 75 80  
 Arg Ala Ser Glu Ala Arg Ala Ala Ala Pro Pro Arg Arg Ala Ser Phe  
 85 90 95  
 Pro Arg Pro Arg Ser Val Thr Ala Arg Ser Gln Ser Val Arg Gly Arg  
 100 105 110  
 40 Arg Asp Ser Ala Ile Thr Arg Ala Pro Arg Gly Gly Tyr Leu Gly Pro  
 115 120 125  
 Met Asp Pro Arg Asp Val Leu Gly Arg Val Gly Gly Ser Arg Val Val  
 130 135 140

Pro Ser Pro Leu Phe Leu Asp Glu Leu Ser Tyr Glu Glu Asp Asp Tyr  
 145 150 155 160  
 Pro Ala Ala Val Ala His Asp Asp Gly Ala Gly Ala Arg Pro Pro Ala  
 165 170 175  
 5 Thr Val Glu Ile Ile Glu Gly Arg Val Ser Gly Pro Glu Leu Gln Ala  
 180 185 190  
 Ala Phe Pro Leu Asp Arg Leu Thr Pro Arg Val Ala Ala Trp Asp Glu  
 195 200 205  
 Ser Val Arg Ser Ala Leu Gly His Pro Ala Gly Phe Tyr Pro Cys Pro  
 10 210 215 220  
 Asp Ser Ala Phe Gly Leu Ser Arg Val Gly Val Met His Phe Asp Ala  
 225 230 235 240  
 Asp Pro Lys Val Phe Phe Arg Gln Thr Leu Gln Gln Gly Glu Ala Trp  
 245 250 255  
 15 Tyr Val Thr Gly Asp Ala Ile Leu Asp Leu Thr Asp Arg Arg Ala Lys  
 260 265 270  
 Thr Ser Pro Ser Arg Ala Met Gly Phe Leu Val Asp Ala Ile Val Arg  
 275 280 285  
 Val Ala Ile Asn Gly Trp Val Cys Gly Thr Arg Leu His Thr Glu Gly  
 20 290 295 300  
 Arg Gly Ser Glu Leu Asp Asp Arg Ala Ala Glu Leu Arg Arg Gln Phe  
 305 310 315 320  
 Ala Ser Leu Thr Ala Leu Arg Pro Val Gly Ala Ala Ala Val Pro Leu  
 325 330 335  
 25 Leu Ser Ala Gly Gly Ala Ala Pro Pro His Pro Gly Pro Asp Ala Ala  
 340 345 350  
 Val Phe Arg Ser Ser Leu Gly Ser Leu Leu Tyr Trp Pro Gly Val Arg  
 355 360 365  
 Ala Leu Leu Gly Arg Asp Cys Arg Val Ala Ala Arg Tyr Ala Gly Arg  
 30 370 375 380  
 Met Thr Tyr Ile Ala Thr Gly Ala Leu Leu Ala Arg Phe Asn Pro Gly  
 385 390 395 400  
 Ala Val Lys Cys Val Leu Pro Arg Glu Ala Ala Phe Ala Gly Arg Val  
 405 410 415  
 35 Leu Asp Val Leu Ala Val Leu Ala Glu Gln Thr Val Gln Trp Leu Ser  
 420 425 430  
 Val Val Val Gly Ala Arg Leu His Pro His Ser Ala His Pro Ala Phe  
 435 440 445  
 Ala Asp Val Glu Gln Glu Ala Leu Phe Arg Ala Leu Pro Leu Gly Ser  
 40 450 455 460  
 Pro Gly Val Val Ala Ala Glu His Glu Ala Leu Gly Asp Thr Ala Ala  
 465 470 475 480  
 Arg Arg Leu Leu Ala Thr Ser Gln Ala Val Leu Gly Ala Ala Val Tyr

```

                485                490                495
Ala Leu His Thr Ala Thr Val Thr Leu Lys Tyr Ala Cys Gly Asp Ala
                500                505                510
Arg Arg Arg Arg Asp Asn Ala Ala Ala Ala Arg Ala Val Leu Ala Thr
5   515                520                525
Gly Leu Ile Leu Gln Arg Leu Leu Gly Leu Ala Asp Thr Val Val Ala
    530                535                540
Cys Val Ala Ala Phe Asp Gly Gly Ser Thr Ala Pro Glu Val Gly Thr
545                550                555                560
10  Tyr Thr Pro Leu Arg Tyr Ala Cys Val Leu Arg Ala Thr Gln Pro Leu
    565                570                575
Tyr Ala Arg Thr Thr Pro Ala Lys Phe Trp Ala Asp Val Arg Ala Ala
    580                585                590
Ala Glu His Val Asp Leu Arg Pro Ala Ser Ser Ala Pro Arg Ala Pro
15  595                600                605
Val Ser Gly Thr Ala Asp Pro Ala Phe Leu Leu Glu Asp Leu Ala Ala
    610                615                620
Phe Pro Pro Ala Pro Leu Asn Ser Glu Ser Val Leu Gly Pro Arg Val
625                630                635                640
20  Arg Val Val Asp Ile Met Ala Gln Phe Arg Lys Leu Leu Met Gly Asp
    645                650                655
Glu Glu Thr Ala Ala Leu Arg Ala His Val Ser Gly Arg Arg Ala Thr
    660                665                670
Gly Leu Gly Gly Pro Pro Arg Pro
25  675                680

```

## (2) INFORMATION FOR SEQ ID NO:111:

## (i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 556 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 35 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

```

Val Ile Leu Lys Met Arg Gly Gly Gly Arg Glu Met Ser Val Ile Gly
40  1             5             10             15
Asp Ala Arg His Pro Arg Gln Phe Pro Ser Gln Gly Pro Arg Pro Phe
    20             25             30
Ser Val Ala Gly Pro Gly Ser Leu Pro Pro Ser Pro Pro Gly Ala

```

	35		40		45	
	Arg	Ala	Leu	Leu	Ile	Arg
	50		55		60	
	Ala	Pro	Met	Asp	Leu	Leu
5	65		70		75	80
	Val	Ser	Pro	Pro	Pro	Arg
			85		90	95
	Ala	Ala	Pro	Pro	Leu	Tyr
			100		105	110
10	Met	Pro	Ser	Pro	Pro	Met
			115		120	125
	Leu	Leu	Asp	Asp	Leu	Gly
			130		135	140
	Leu	Asp	Thr	Trp	Asn	Glu
15	145		150		155	160
	Asp	Met	Tyr	Arg	Glu	Cys
			165		170	175
	Ile	Asp	Trp	Gly	Asp	Ala
			180		185	190
20	Arg	Ala	His	Gly	Asp	Val
			195		200	205
	Glu	Leu	Pro	Ser	Tyr	Tyr
			210		215	220
	Leu	Arg	Ala	Arg	Glu	Glu
25	225		230		235	240
	Ser	Ala	Leu	Tyr	Arg	Tyr
			245		250	255
	Gln	Ala	His	Met	Arg	Gly
			260		265	270
30	Thr	Thr	Ile	Ala	Asp	Arg
			275		280	285
	Val	Leu	Phe	Leu	His	Leu
			290		295	300
	Ala	Ala	Tyr	Ala	Glu	Gln
35	305		310		315	320
	Cys	Cys	Asp	Leu	Glu	Ser
			325		330	335
	Leu	Met	Phe	Ile	Asn	Gly
			340		345	350
40	Ala	Arg	Arg	Leu	Arg	Glu
			355		360	365
	Pro	Leu	Val	Arg	Ser	Ala
			370		375	380



Thr Pro Pro Val Leu Gln Gly Asn Gln Ala Arg Ser Ser Gly Tyr Phe  
 385 390 395 400  
 Met Leu Leu Ile Arg Ala Lys Leu Asp Ser Tyr Ser Ser Val Ala Thr  
 405 410 415  
 5 Ser Glu Gly Glu Ser Val Met Arg Glu His Ala Tyr Ser Arg Gly Arg  
 420 425 430  
 Thr Arg Asn Asn Tyr Gly Ser Thr Ile Glu Gly Leu Leu Asp Leu Pro  
 435 440 445  
 Asp Asp Asp Asp Ala Pro Ala Glu Ala Gly Leu Val Ala Pro Arg Met  
 10 450 455 460  
 Ser Phe Leu Ser Ala Gly Gln Arg Pro Arg Arg Leu Ser Thr Thr Ala  
 465 470 475 480  
 Pro Ile Thr Asp Val Ser Leu Gly Asp Glu Leu Arg Leu Asp Gly Glu  
 485 490 495  
 15 Glu Val Asp Met Thr Pro Ala Asp Ala Leu Asp Asp Phe Asp Leu Glu  
 500 505 510  
 Met Leu Gly Asp Val Glu Ser Pro Ser Pro Gly Met Thr His Asp Pro  
 515 520 525  
 Val Leu Tyr Gly Ala Leu Asp Val Asp Asp Phe Glu Phe Glu Gln Met  
 20 530 535 540  
 Phe Thr Asp Ala Met Gly Ile Asp Asp Phe Gly Gly  
 545 550 555

## (2) INFORMATION FOR SEQ ID NO:112:

25

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7362 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

35 CGCGGGGGAG GGGACGACGC GGGGGAGGGG ACGACGCGGG GGAGGGGAGG ACGCGGGGGA 60  
 TATATAAAGC GGTACAAAGC GCGGGAATGG GCATATTGGA CCCGCGTGAT TCGGTTGCTC 120  
 GCGGTTGTCT TGT TTGGACG TTTT TATGC GGAACAAGG GGGCTTACCG GTTACACTGT 180  
 CCGCTCGCTA TGGGTTTCGT CTGTCTGTTT GGGCTTGTCTG TTATGGGAGC CTGGGGGGCG 240  
 TGGGGTGGGT CACAGGCAAC CGAATATGTT CTTCGTAGTG TTATTGCCAA AGAGGTGGGG 300  
 40 GACATACTAA GAGTGCCTTG CATGCGGACC CCCGCGGACG ATGTTTCTTG GCGCTACGAG 360  
 GCCCCGTCCG TTATTGACTA TGCCCGCATA GACGGAATAT TTCTTCGCTA TCACTGCCCCG 420  
 GGGTTGGACA CGTTTTTGTG GGATAGGCAC GCCCAGAGGG CGTATCTGGT TAACCCCTTT 480  
 CTCTTTGCGG CGGGATTTT GGAGGACTTG AGTCACTCTG TGTTCGGC CGACACCCAG 540

	GAAACAACGA	CGCGCCGGGC	CCTTTATAAA	GAGATACGCG	ATGCGTTGGG	CAGTCGAAAA	600
	CAGGCCGTCA	GCCACGCACC	CGTCAGGGCC	GGGTGTGTAA	ACTTTGACTA	CTCACGCACT	660
	CGCCGCTGCG	TCGGGCGACG	CGATTTACGG	CCTGCCAACA	CCACGTCAAC	GTGGGAACCG	720
	CCTGTGTCGT	CGGACGATGA	AGCGAGCTCG	CAGTCGAAGC	CCCTCGCCAC	CCAGCCGCC	780
5	GTCTTCGCCC	TTTCGAACGC	CCCCCACC	CGGGTCTCCC	CGACGCGAGG	TCGGCGCCGG	840
	CATACTCGCC	TCCGACGCAA	CTAGCCACGT	CTGCATCGCA	AGCCACCCTG	GGTCGGGAGC	900
	AGGATATCCG	ACCCGTCTAG	CGGCCGGGTC	GGCTGTCCAG	CGTCGTGCGC	CTAGAGGCTG	960
	TCCGCCGGGC	GTGATGTTTT	CCGCATCTAC	GACCCCCGAA	CAGCCCCTGG	GGCTGTCCGG	1020
	CGATGCGACG	CCGCCCTG	CGACTTCCGT	GCCCCTGGAC	TGGGCCGCGT	TTCGGCGCGC	1080
10	GTTTCTGATC	GACGACGCCT	GGCGGCCCTT	GTTGGAGCCG	GAGCTCGCGA	ACCCCTAAC	1140
	CGCGCGCCTC	CTCGCGGAGT	ATGACCGTCG	GTGCCAGACC	GAAGAGGTGC	TGCCGCCGCG	1200
	GGAGGATGTG	TTCTCCTGGA	CGCGGTATTG	TACCCCCGAC	GACGTGCGCG	TGGTTATCAT	1260
	CGGGCAGGAC	CCGTACCACC	ATCCCGGCCA	GGCGCACGGC	CTGGCGTTTA	GCGTGCGTGC	1320
	GGATGTGCCG	GTGCCTCCGA	GTCTACGGAA	CGTGCTGGCG	GCGGTAAAAA	ATTGTTACCC	1380
15	CGACGCGCGC	ATGAGCGGCC	GCGGCTGCCT	GGAAAAGTGG	GCTCGCGACG	GCGTGCTGTT	1440
	GTTGAACACG	ACCTTGACCG	TCAAGCGCGG	GGCGCGCGCG	TCCCACTCCA	AGCTTGGATG	1500
	GGACCCTTTT	GTGGGCGGGG	TGGTCCGACG	GCTGGCCGCG	CGCCGCCCGG	GCCTGGTCTT	1560
	TATGCTCTGG	GGCGCCCATG	CCCAGAACGC	GATCAGGCC	GACCTCGCC	AACACTACGT	1620
	CCTCAAGTTT	TCTCACCCGT	CGCCCCCTC	CAAGGTCCCG	TTTGGGACGT	GCCAGCATTT	1680
20	CCTCGCCGCG	AATCGCTACC	TCGAAACCCG	GGACATTATG	CCTATCGACT	GGTCGGTATA	1740
	AGATGCCGAC	ATCCGGGGTC	TTGATTTACG	AGGGGGCAAT	TAATAAAGAC	TGTTGATGGT	1800
	TAAATCTCGG	GTCTCATACC	GGTCCGTGAT	GTCCGGCGTG	GGGGAAGAGA	GGGTCCCTC	1860
	TGCGTTTACT	ATCCTTGCCCT	CGTGGGGCTG	GACGTTTGCA	CCCCAGAACC	ATGATCCTGG	1920
	CGCGTCGCGG	AATACGACGC	CCATAGAGTC	GATTGCGGGG	ACCGCACCGG	ACGCGCACGT	1980
25	GGGGCCTCTC	GACGGAGAGC	CGGACCGGGA	TGCGATCTCC	CCGCTTACGT	CGAGCGTGGC	2040
	CGGCGACCCG	CCGGGGGCGG	ACGGCCCCCTA	CGTCACCTTT	GATACTCTGT	TTATGGTATC	2100
	TTCGATCGAC	GAAGTGGGGC	GCCGCCAGCT	CACGGATACG	ATCCGTAAGG	ACCTGCGGCT	2160
	GTGCTGGCC	AAGTTCAGCA	TCGCGTGATC	CAAGACCTCG	TCGTTTTTCG	GGACGGCCGC	2220
	GCGCCAGCGC	AAGCGCGGAG	CACCGCCGCA	ACGCACATGC	GTACCACGCA	GCAACAAGAG	2280
30	CCTCCAGATG	TTGTTTTTGT	GCAAGCGCGC	CAACGCCGCG	CAGGTGCGCG	AGCAGCTGCG	2340
	GGCGGTTATT	CGGTGCGGCA	AGCCGCGCAA	GTATTACACG	CGGTCTTCGG	ATGGGCGGCT	2400
	CTGCCCGGCC	GTCCCGTGT	TTGTACACGA	GTTTGTTCG	TCCGAACCCA	TGCGCCTCCA	2460
	TCGAGATAAC	GTCATGCTGT	CTACGGAACC	AGACTAAGCA	CCCCCGCCGT	CCCCTTTCTT	2520
	TTCCCCCTAC	CCTTCCCCC	GTTACTGATG	TGTTGTGACG	TTTCAATAAA	TAACACGTAG	2580
35	CTTATTTTGT	TGGATGATGG	ATTGATTGAT	TTTATTGACC	GTTGCTTCGC	CCGGCGGTGC	2640
	CGTCGCCGCG	CGCAGAGGGA	ATATGCAAGC	GGCGGGGGTG	GGGAGGAAAG	AAGGTTTCAG	2700
	GTTCCGGGGG	TTGGGTCTGC	GTGCTCCAGG	GTGGGGCTGA	TCTGAATTTC	CCGCAGAACC	2760
	TCGACCAGTA	GGTCTGTTGT	GTTTGTGTTG	AACTCGCCCG	CCGTTGGGGA	TACGGGGGCG	2820
	GGGGGTGTGG	TTGGGCGGAC	GTCCAGGGGT	GCCTTATCGC	ACCCCGCGCG	CGCCTCGGGG	2880
40	GCCGTCCCGT	AGATCGTTGC	GGTGATGTAG	ATGGTGTCG	GGGTCCACAC	CACCGTCAGG	2940
	ATGCCGGCCG	TCGCACTCCG	GACGCTTTCG	CCGTGCGATG	AGCTGACCCA	GGAGTCAAAG	3000
	GGGTACGCGT	ACATATGGGC	GTCCCAACCAG	CGCTCCAGCC	TCTGGGTACT	AGCGCGTCTT	3060
	ATAAAGCGGT	ATGCGCAAAA	TTGCGCACGA	CAGTCGATAA	TCACCAGCAG	CCCGATGGGG	3120

	GTGTGTTGTA	TCACCACGCC	TCCGCGGGGC	AGGCGGTCCT	GGCGCGCTCG	ACCCCGCGTC	3180
	AGAACCGCGC	GCGTCCCTGA	CTCAAACACG	TGCACCACCT	GTGCCGCGTC	CGGCAGCGCG	3240
	CTCGTTAGCG	ACGCCCTGGG	GTGATGTAGG	CTGTACGCGA	TGGTCGTCTG	GGGGTTCCCC	3300
	ATGTCTCGGG	GGGGTGGGGG	TGAATGTCAC	CCGCCCCGGG	TGCGGTGGGA	ACGCGAGGGA	3360
5	ATGGAGGGTT	AATAGACAAT	GACCACATTC	GGATCGCGTA	GAGCAGATAG	TATGTGCTCG	3420
	CTAATGACGT	CATCGCGTTC	GTGGCGCTCC	CGGAGCGGGT	TTAGATTCAT	GTGCAGGAAC	3480
	TCGGATGAGG	TGGTGCGGGA	CATGGCTACG	TACGCGCTGT	TTAGGCGCAG	GTTTCCGGGC	3540
	GTGAAGCATA	TGGCGACCTT	GTCCAGACTG	AGCCCCCTGG	AGCGCGTGAT	GGTCATCGCG	3600
	AGTTTGGAGC	TGATGCCGTA	GTCCGCGTTG	ATGCCCATGG	CCAGCTCCGT	GGAGTCGATC	3660
10	GACTCGACAA	ACTCACTGAT	GTTGGTATTG	ACGACAGACA	TGAAGCCGTG	CTGGTCCCGC	3720
	AGGACGATGT	AGGGCAGGGG	GGACTCCTCC	AAGAACTCGG	CCACGCCCGC	CGTCGCGTGC	3780
	CGCCGCCGCA	GCTCCTCCGC	GAACGCGAAC	ACCCGGGTGT	ACGTGTACCC	CATCAGCGTG	3840
	TAGTTGTCCG	TCTGCAGGGC	CACGGACATC	AGCCCCCGCG	GCGGCGAGCC	GGTCAGCAGC	3900
	TCGACAGCCCC	GGAAGATGAC	ATTGTCCACG	TAGGTGCTGA	AGGGGGCGCT	CTCAAACACC	3960
15	TCCCCGAAGA	GCTCCCGTAG	GATAAGGTAT	CGCCCCAGAA	AGGCCCTCTT	CAGGAGCCCA	4020
	AACTGGGCGT	GGACGGCCGC	GGTGGTCTCC	GGCTCTTCGA	GGGCGTAGTG	GCAGTAGAAC	4080
	ACGTCCAGCT	GCTGTTCGTC	CAGCCCCGCG	AAGATAACGT	CAAGGTCGTC	GTCGGGGAAG	4140
	TCGTCCGGGC	CCCCGTCCCG	CGGGCCCAGG	TGCTTAAAT	TGAACGCACG	CTCCCCCGGA	4200
	GAGCGGTGCG	TGGTGTCCGC	GGCCCTGGTT	GCCGATGCGC	CGGCGGCGTC	CCGGCGTAGC	4260
20	GACAGGAGTT	CTGCCGTCAG	CTCCCCTAGG	CGGCCGTAGG	CCAGGGTCCT	CTGGGTCCGC	4320
	TCCAGGCCGG	GGCGCTGGAG	AAAGTTGTAA	AAGTGAATCA	GCCCGCCGAA	CATGAGCCGC	4380
	GACAGGAACC	GGTAGGCGAA	CTCCACCGAG	GTCTCCCCCT	GGGTCTTCAC	GAAGCTGTCT	4440
	TCGCGCAGCA	CAGCCTCGAA	GGTCCGAAAC	GTCCCGTCGA	ACCCAAACAC	CATCTTTCGG	4500
	AGGCGCGCGG	TCACCGCGAC	CTGGCTGTTG	AGGACGTACG	TGATGTCTGT	CCGGGCCACG	4560
25	ACTAGCTGTT	GCTTGCTGTG	CACCTCACAG	CGCACGTGCC	CCGCGTCCTG	GTCTGACTC	4620
	TGGGAGTAGT	TGGTGATGCG	ACTGGCGTTG	GCCGTGATCC	ACTTTTCCAT	GGTCAGCGTG	4680
	GGTTGCTGCG	TGAGCCGTCG	ATACTCGTCA	AACTCTTTGA	CCGACACAAA	CGTAAGCACG	4740
	GGGAGGGTAA	ACACAACAAA	CTCCCCCTCG	CGAGTCACCT	TTAGGTAGGC	GTGGAGCTTG	4800
	GCCATGTACG	CGCTGACCTC	CTTGTTGGGAC	GAGAACAGCC	GCGTCCACCC	CGGAAGGTTG	4860
30	GCCGGGTTGG	TGATGTAAC	TTCCGGGACG	ACAAAGCGGT	CCACAAACTG	CATGTGCTCC	4920
	TCGGTGATGG	GAAGGCCGTA	CTCCAGCACC	TTCATGAGGT	TCCCGAACTC	GTGCTCCACA	4980
	CATCGCTTGT	TGTTAATGAA	AATGGCCCGG	CTGTGCGAGA	GGCGCGTGTA	CTCGCGTAGG	5040
	GTGCGGTTGC	AGATGAGGTA	CGTGAGCAGC	TTTTCGCTCT	GCCGGACGGA	GCATCGCAGT	5100
	TTTTGGTGTT	CGAAGGTGGA	CTCCAGCGAG	GCCGTCTGGG	TCGGCGACCC	CACGCACACC	5160
35	AGCACCGGCC	GCAGGCGGCC	CGCGTACTGG	GGGGTGTGGT	ACAGGGCGTT	AATCATCCAC	5220
	CAGCAATACA	CCACGGTCGT	GAGTAGGTGC	CGCCCCAGGA	GCCCCGCCCT	GTCGATGACG	5280
	ATAATGTTGC	TGCGGGTGAA	AGCCGCGCAG	GCCCCGTGTG	TGACCGAGGC	CAGGCGCGTG	5340
	AGGGCACCCCT	GGCCCAGCCC	CAAAGTCTGC	TCTAGGGCGG	TGAGGGCGTG	GAACCTGTTT	5400
	CGCGCGTCTT	CGCCCCCGTG	CGCCGCCAGG	GCCCGCTTGG	TGATGTCGAG	GATCACCTCC	5460
40	CAGTAGTACG	TCAGGTCTCG	CCGCTGCAGG	TCTTCCAGCG	AGGCGGGGCT	GCTGCCCAGG	5520
	GTGTACGGGT	GCTGCCCCAG	CTGGGCCTGG	ACGTGATTCC	CGCGAAACCC	GAACCTCGTA	5580
	AAGATGGTGT	TGATGGGTCG	ACTCAGAAAC	GCCCCCGAGA	GCTTAACGTA	CATGTTCTGC	5640
	GCCGCGATTTC	GCGTGGCGCC	CGTGACCACG	CAGTCCAGGA	CCTCGTTGAG	GGTCTGCACG	5700

CACGTACTCT TTCCGGATCC GCGGTGCGG GTGATGAGAT ACGCCGCGAA CGGAAACTCC 5760  
 CGGAGCGGCA GGCCGGTCGG GACCTCCAAG GCCGCCACGT CCCGGAACCA CTGCAGGCGC 5820  
 GGCACCTGCG TGACGTCGAG CTGCTGCTGC GAGAGCTCTC GGATGCGTGC GATGATTGGT 5880  
 TGGACCCCGT GCATGGACGT AAAATTAAAA AACGCCTCGT CCCTGAACCG CACGGCGGGT 5940  
 5 CTGGCCCCGG GCTGCTGTGG GGGCGGACCT GGTGCCCGGA CGTCCCGCGA GCCCTCCCCG 6000  
 CCGGACGCCG CCATGGCCGC ACAGCGCGCG CGGGCGCCGG CGATGCGGAC GCGGGGCGGC 6060  
 GACGCGGCGC TATGCGCCCC CGAGGACGGC TGGGTGAAGG TTCACCCAC CCCC GGACG 6120  
 ATGTTGTTCC GCGAGATTCT CCTCGGGCAG ATGGGGTACA CCGAGGGTCA GGGGGTGTAC 6180  
 AACGTCGTCC GGTCCAGCGA GGCCGCCACC CGACAGCTGC AGGCGGCGAT CTTCCACGCG 6240  
 10 CTCTCAACG CCACAACGTA CCGGGACCTG GAGGAGGACT GGCGCCGCCA CGTGGTGGCC 6300  
 CGCGGCCTCC AGCCGACGCG GCTGGTTCGC AGGTACCGGA ACGCCCGGA GGGCGATATC 6360  
 GCCGGGGTGG CCGAGCGGGT GTTCGACACG TGGCGATGCA CGCTCAGGAC GACGCTGCTG 6420  
 GACTTTGCCC ACGGGGTGGT AGACTGCTTT GCGCCGGGCG GCCCAAGCG ACCGACCAGC 6480  
 TTCCCCAAT ATATCGACTG GCTGACGTGT CTGGGGCTGG TTCCCATATT GCGCAAGACG 6540  
 15 CGCGAGGGGG AGGCGACGCA GCGCCTGGGG GCGTTTCTCA GGCAGCACAC GCTGCCCCGG 6600  
 CAGCTGGCCA CGGTCGCCGG GGCCGCGGAG CGCGCCGGCC CGGGGCTTCT GGAGCTGGCC 6660  
 GTCGCGTTCG ACTCCACGCG CATGGCGGAA TACGACCGTG TGCACATCTA CTACAACCAT 6720  
 CGCCGGGGGG AGTGGCTGGT GCGCGACCCG GTCAGCGGGC AGCGCGGCGA GTGCCTGGTG 6780  
 CTGTGCCCCC CCCTGTGGAC CGGCGACCGC CTGGTCTTCG ATTCGCCCCG TCAGCGGCTG 6840  
 20 TGCCCCGAGA TCGTCGCGTG CCACGCCCTC CGGGAACACG CGCACATCTG CCGTCTGCGC 6900  
 AACACCGCGT CCGTCAAGGT GCTGTTGGGG CGCAAGAGCG ACAGCGAGCG CGGGGTGGCT 6960  
 GGCGCCGCGC GGGTCGTCAA TAAGGCGCTG GGGGAGGATG ACGAGACGAA GGCCGGCTCG 7020  
 GCCGCTCGC GTCTCGTGC GCTCATCATC AACATGAAGG GCATGCGCCA CGTGGGCGAC 7080  
 ATCAACGACA CGGTACGCGC CTACTTGGAC GAGCGGGGG GGCACCTGAT CGACACCCCC 7140  
 25 GCCGTCGACC ACACCTCCC TGGGTTCCGGC AAGGGCGGCA CCGGCCCGCG GTCGGCGGCC 7200  
 CAGGACCCGG GGGCGCGACC GCAGCAGCTT CGCCAGGCGT TTCAGACGGC CGTGGTCAAC 7260  
 AACATCAACG GCATGCTGGA GGGCTATATC AATAATCTCT TTGGAACCAT AGAACGCTG 7320  
 CGAGAGACGA ACGCGGGTCT GGCGACCCAG CTGCAGGCGC G 7362

30 (2) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 180 amino acids

(B) TYPE: amino acid

35 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

Met Arg Thr Pro Ala Asp Asp Val Ser Trp Arg Tyr Glu Ala Pro Ser

1

5

10

15

Val Ile Asp Tyr Ala Arg Ile Asp Gly Ile Phe Leu Arg Tyr His Cys  
                   20                                  25                                  30  
 Pro Gly Leu Asp Thr Phe Leu Trp Asp Arg His Ala Gln Arg Ala Tyr  
                   35                                  40                                  45  
 5 Leu Val Asn Pro Phe Leu Phe Ala Ala Gly Phe Leu Glu Asp Leu Ser  
                   50                                  55                                  60  
 His Ser Val Phe Pro Ala Asp Thr Gln Glu Thr Thr Thr Arg Arg Ala  
                   65                                  70                                  75                                  80  
 Leu Tyr Lys Glu Ile Arg Asp Ala Leu Gly Ser Arg Lys Gln Ala Val  
                                   85                                  90                                  95  
 10 Ser His Ala Pro Val Arg Ala Gly Cys Val Asn Phe Asp Tyr Ser Arg  
                                   100                                  105                                  110  
 Thr Arg Arg Cys Val Gly Arg Arg Asp Leu Arg Pro Ala Asn Thr Thr  
                   115                                  120                                  125  
 15 Ser Thr Trp Glu Pro Pro Val Ser Ser Asp Asp Glu Ala Ser Ser Gln  
                   130                                  135                                  140  
 Ser Lys Pro Leu Ala Thr Gln Pro Pro Val Leu Ala Leu Ser Asn Ala  
                   145                                  150                                  155                                  160  
 Pro Pro Arg Arg Val Ser Pro Thr Arg Gly Arg Arg Arg His Thr Arg  
                                   165                                  170                                  175  
 20 Leu Arg Arg Asn  
                                   180

## (2) INFORMATION FOR SEQ ID NO:114:

25

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 334 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

30

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

35

Met Lys Arg Ala Arg Ser Arg Ser Pro Ser Pro Pro Ser Arg Pro Ser  
   1                  5                                  10                                  15  
 Ser Pro Phe Arg Thr Pro Pro His Gly Gly Ser Pro Arg Arg Glu Val  
                   20                                  25                                  30  
 40 Gly Ala Gly Ile Leu Ala Ser Asp Ala Thr Ser His Val Cys Ile Ala  
                   35                                  40                                  45  
 Ser His Pro Gly Ser Gly Ala Gly Tyr Pro Thr Arg Leu Ala Ala Gly  
                   50                                  55                                  60

Ser Ala Val Gln Arg Arg Arg Pro Arg Gly Cys Pro Pro Gly Val Met  
 65 70 75 80  
 Phe Ser Ala Ser Thr Thr Pro Glu Gln Pro Leu Gly Leu Ser Gly Asp  
 85 90 95  
 5 Ala Thr Pro Pro Leu Pro Thr Ser Val Pro Leu Asp Trp Ala Ala Phe  
 100 105 110  
 Arg Arg Ala Phe Leu Ile Asp Asp Ala Trp Arg Pro Leu Leu Glu Pro  
 115 120 125  
 Glu Leu Ala Asn Pro Leu Thr Ala Arg Leu Leu Ala Glu Tyr Asp Arg  
 10 130 135 140  
 Arg Cys Gln Thr Glu Glu Val Leu Pro Pro Arg Glu Asp Val Phe Ser  
 145 150 155 160  
 Trp Thr Arg Tyr Cys Thr Pro Asp Asp Val Arg Val Val Ile Ile Gly  
 165 170 175  
 15 Gln Asp Pro Tyr His His Pro Gly Gln Ala His Gly Leu Ala Phe Ser  
 180 185 190  
 Val Arg Ala Asp Val Pro Val Pro Pro Ser Leu Arg Asn Val Leu Ala  
 195 200 205  
 Ala Val Lys Asn Cys Tyr Pro Asp Ala Arg Met Ser Gly Arg Gly Cys  
 20 210 215 220  
 Leu Glu Lys Trp Ala Arg Asp Gly Val Leu Leu Leu Asn Thr Thr Leu  
 225 230 235 240  
 Thr Val Lys Arg Gly Ala Ala Ala Ser His Ser Lys Leu Gly Trp Asp  
 245 250 255  
 25 Arg Phe Val Gly Gly Val Val Arg Arg Leu Ala Ala Arg Arg Pro Gly  
 260 265 270  
 Leu Val Phe Met Leu Trp Gly Ala His Ala Gln Asn Ala Ile Arg Pro  
 275 280 285  
 Asp Pro Arg Gln His Tyr Val Leu Lys Phe Ser His Pro Ser Pro Leu  
 30 290 295 300  
 Ser Lys Val Pro Phe Gly Thr Cys Gln His Phe Leu Ala Ala Asn Arg  
 305 310 315 320  
 Tyr Leu Glu Thr Arg Asp Ile Met Pro Ile Asp Trp Ser Val  
 325 330

35

(2) INFORMATION FOR SEQ ID NO:115:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 231 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

```

5  Met Val Lys Ser Arg Val Ser Tyr Arg Ser Val Met Ser Gly Val Gly
   1           5           10           15
   Glu Glu Arg Val Pro Ser Ala Phe Thr Ile Leu Ala Ser Trp Gly Trp
      20           25           30
   Thr Phe Ala Pro Gln Asn His Asp Pro Gly Asp Asn Thr Thr Pro Ile
10      35           40           45
   Glu Ser Ile Ala Gly Thr Ala Pro Asp Ala His Val Gly Pro Leu Asp
      50           55           60
   Gly Glu Pro Asp Arg Asp Ala Ile Ser Pro Leu Thr Ser Ser Val Ala
      65           70           75           80
15  Gly Asp Pro Pro Gly Ala Asp Gly Pro Tyr Val Thr Phe Asp Thr Leu
      85           90           95
   Phe Met Val Ser Ser Ile Asp Glu Leu Gly Arg Arg Gln Leu Thr Asp
      100          105          110
   Thr Ile Arg Lys Asp Leu Arg Leu Ser Leu Ala Lys Phe Ser Ile Ala
20      115          120          125
   Cys Thr Lys Thr Ser Ser Phe Ser Gly Thr Ala Ala Arg Gln Arg Lys
      130          135          140
   Arg Gly Ala Pro Pro Gln Arg Thr Cys Val Pro Arg Ser Asn Lys Ser
      145          150          155          160
25  Leu Gln Met Phe Val Leu Cys Lys Arg Ala Asn Ala Ala Gln Val Arg
      165          170          175
   Glu Gln Leu Arg Ala Val Ile Arg Ser Arg Lys Pro Arg Lys Tyr Tyr
      180          185          190
   Thr Arg Ser Ser Asp Gly Arg Leu Cys Pro Ala Val Pro Val Phe Val
30      195          200          205
   His Glu Phe Val Ser Ser Glu Pro Met Arg Leu His Arg Asp Asn Val
      210          215          220
   Met Leu Ser Thr Glu Pro Asp
      225          230
35

```

(2) INFORMATION FOR SEQ ID NO:116:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 199 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

```

5  Met Gly Asn Pro Gln Thr Thr Ile Ala Tyr Ser Leu His His Pro Arg
   1           5           10           15
   Ala Ser Leu Thr Ser Ala Leu Pro Asp Ala Ala Gln Val Val His Val
     20           25           30
   Phe Glu Ser Gly Thr Arg Ala Val Leu Thr Arg Gly Arg Ala Arg Gln
10      35           40           45
   Asp Arg Leu Pro Arg Gly Gly Val Val Ile Gln His Thr Pro Ile Gly
     50           55           60
   Leu Leu Val Ile Ile Asp Cys Arg Ala Glu Phe Cys Ala Tyr Arg Phe
   65           70           75           80
15  Ile Gly Arg Ala Ser Thr Gln Arg Leu Glu Arg Trp Trp Asp Ala His
     85           90           95
   Met Tyr Ala Tyr Pro Phe Asp Ser Trp Val Ser Ser Ser His Gly Glu
     100          105          110
   Ser Val Arg Ser Ala Thr Ala Gly Ile Leu Thr Val Val Trp Thr Pro
20      115          120          125
   Asp Thr Ile Tyr Ile Thr Ala Thr Ile Tyr Gly Thr Ala Pro Glu Ala
     130          135          140
   Arg Cys Asp Asn Ala Pro Leu Asp Val Arg Pro Thr Thr Pro Pro Ala
   145          150          155          160
25  Pro Val Ser Pro Thr Ala Gly Glu Phe Pro Ala Asn Thr Thr Asp Leu
     165          170          175
   Leu Val Glu Val Leu Arg Glu Ile Gln Ile Ser Pro Thr Leu Asp Asp
     180          185          190
   Ala Asp Pro Thr Pro Gly Thr
30      195

```

(2) INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:

```

35  (A) LENGTH: 877 amino acids
     (B) TYPE: amino acid
     (C) STRANDEDNESS: single
     (D) TOPOLOGY: linear

```

40 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:



Met Ala Ala Ser Gly Gly Glu Gly Ser Arg Asp Val Arg Ala Pro Gly  
 1 5 10 15  
 Pro Pro Pro Gln Gln Pro Gly Ala Arg Pro Ala Val Arg Phe Arg Asp  
 20 25 30  
 5 Glu Ala Phe Leu Asn Phe Thr Ser Met His Gly Val Gln Pro Ile Ile  
 35 40 45  
 Ala Arg Ile Arg Glu Leu Ser Gln Gln Gln Leu Asp Val Thr Gln Val  
 50 55 60  
 Pro Arg Leu Gln Trp Phe Arg Asp Val Ala Ala Leu Glu Val Pro Thr  
 10 65 70 75 80  
 Gly Leu Pro Leu Arg Glu Phe Pro Phe Ala Ala Tyr Leu Ile Thr Gly  
 85 90 95  
 Asn Ala Gly Ser Gly Lys Ser Thr Cys Val Gln Thr Leu Asn Glu Val  
 100 105 110  
 15 Leu Asp Cys Val Val Thr Gly Ala Thr Arg Ile Ala Ala Gln Asn Met  
 115 120 125  
 Tyr Val Lys Leu Ser Gly Ala Phe Leu Ser Arg Pro Ile Asn Thr Ile  
 130 135 140  
 Phe His Glu Phe Gly Phe Arg Gly Asn His Val Gln Ala Gln Leu Gly  
 20 145 150 155 160  
 Gln His Pro Tyr Thr Leu Ala Ser Ser Pro Ala Ser Leu Glu Asp Leu  
 165 170 175  
 Gln Arg Arg Asp Leu Thr Tyr Tyr Trp Glu Val Ile Leu Asp Ile Thr  
 180 185 190  
 25 Lys Arg Ala Ala His Gly Gly Glu Asp Ala Arg Asn Glu Phe His Ala  
 195 200 205  
 Leu Thr Ala Leu Glu Gln Thr Leu Gly Leu Gly Gln Gly Ala Leu Thr  
 210 215 220  
 Arg Leu Ala Ser Val Thr His Gly Ala Leu Pro Ala Phe Thr Arg Ser  
 30 225 230 235 240  
 Asn Ile Ile Val Ile Asp Glu Ala Gly Leu Leu Gly Arg His Leu Leu  
 245 250 255  
 Thr Thr Val Val Tyr Cys Trp Trp Met Ile Asn Ala Leu Tyr His Thr  
 260 265 270  
 35 Pro Gln Tyr Ala Gly Arg Leu Arg Pro Val Leu Val Cys Val Gly Ser  
 275 280 285  
 Pro Thr Gln Thr Ala Ser Leu Glu Ser Thr Phe Glu His Gln Lys Leu  
 290 295 300  
 Arg Cys Ser Val Arg Gln Ser Glu Asn Val Leu Thr Tyr Leu Ile Cys  
 40 305 310 315 320  
 Asn Arg Thr Leu Arg Glu Tyr Thr Arg Leu Ser His Ser Trp Ala Ile  
 325 330 335  
 Phe Ile Asn Asn Lys Arg Cys Val Glu His Glu Phe Gly Asn Leu Met

	340		345		350
	Lys Val Leu Glu Tyr Gly Leu Pro Ile Thr Glu Glu His Met Gln Phe				
	355		360		365
	Val Asp Arg Phe Val Val Pro Glu Ser Tyr Ile Thr Asn Pro Ala Asn				
5	370		375		380
	Leu Pro Gly Trp Thr Arg Leu Phe Ser Ser His Lys Glu Val Ser Ala				
	385		390		395
	Tyr Met Ala Lys Leu His Ala Tyr Leu Lys Val Thr Arg Glu Gly Glu				
	405		410		415
10	Phe Val Val Phe Thr Leu Pro Val Leu Thr Phe Val Ser Val Lys Glu				
	420		425		430
	Phe Asp Glu Tyr Arg Arg Leu Thr Gln Gln Pro Thr Leu Thr Met Glu				
	435		440		445
	Lys Trp Ile Thr Ala Asn Ala Ser Arg Ile Thr Asn Tyr Ser Gln Ser				
15	450		455		460
	Gln Asp Gln Asp Ala Gly His Val Arg Cys Glu Val His Ser Lys Gln				
	465		470		475
	Gln Leu Val Val Ala Arg Asn Asp Ile Thr Tyr Val Leu Asn Ser Gln				
	485		490		495
20	Val Ala Val Thr Ala Arg Leu Arg Lys Met Val Phe Gly Phe Asp Gly				
	500		505		510
	Thr Phe Arg Thr Phe Glu Ala Val Leu Arg Asp Asp Ser Phe Val Lys				
	515		520		525
	Thr Gln Gly Glu Thr Ser Val Glu Phe Ala Tyr Arg Phe Leu Ser Arg				
25	530		535		540
	Leu Met Phe Gly Gly Leu Ile His Phe Tyr Asn Phe Leu Gln Arg Pro				
	545		550		555
	Gly Leu Asp Ala Thr Gln Arg Thr Leu Ala Tyr Gly Arg Leu Gly Glu				
	565		570		575
30	Leu Thr Ala Glu Leu Leu Ser Leu Arg Arg Asp Ala Ala Gly Ala Ser				
	580		585		590
	Ala Thr Arg Ala Ala Asp Thr Ser Asp Arg Ser Pro Gly Glu Arg Ala				
	595		600		605
	Phe Asn Phe Lys His Leu Gly Pro Arg Asp Gly Gly Pro Asp Asp Phe				
35	610		615		620
	Pro Asp Asp Asp Leu Asp Val Ile Phe Ala Gly Leu Asp Glu Gln Gln				
	625		630		635
	Leu Asp Val Phe Tyr Cys His Tyr Ala Leu Glu Glu Pro Glu Thr Thr				
	645		650		655
40	Ala Ala Val His Ala Gln Phe Gly Leu Leu Lys Arg Ala Phe Leu Gly				
	660		665		670
	Arg Tyr Leu Ile Leu Arg Glu Leu Phe Gly Glu Val Phe Glu Ser Ala				
	675		680		685

Pro Phe Ser Thr Tyr Val Asp Asn Val Ile Phe Arg Gly Cys Glu Leu  
 690 695 700  
 Leu Thr Gly Ser Pro Arg Gly Gly Leu Met Ser Val Gln Thr Asp Asn  
 705 710 715 720  
 5 Tyr Thr Leu Met Gly Tyr Thr Tyr Thr Arg Val Phe Ala Phe Ala Glu  
 725 730 735  
 Glu Leu Arg Arg Arg His Ala Thr Ala Gly Val Ala Glu Phe Leu Glu  
 740 745 750  
 Glu Ser Pro Leu Pro Tyr Ile Val Leu Arg Asp Gln His Gly Phe Met  
 10 755 760 765  
 Ser Val Val Asn Thr Asn Ile Ser Glu Phe Val Glu Ser Ile Asp Ser  
 770 775 780  
 Thr Glu Leu Ala Met Ala Ile Asn Ala Asp Tyr Gly Ile Ser Ser Lys  
 785 790 795 800  
 15 Leu Ala Met Thr Ile Thr Arg Ser Gln Gly Leu Ser Leu Asp Lys Val  
 805 810 815  
 Ala Ile Cys Phe Thr Pro Gly Asn Leu Arg Leu Asn Ser Ala Tyr Val  
 820 825 830  
 Ala Met Ser Arg Thr Thr Ser Ser Glu Phe Leu His Met Asn Leu Asn  
 20 835 840 845  
 Pro Leu Arg Glu Arg His Glu Arg Asp Asp Val Ile Ser Glu His Ile  
 850 855 860  
 Leu Ser Ala Leu Arg Asp Pro Asn Val Val Ile Val Tyr  
 865 870 875

25

## (2) INFORMATION FOR SEQ ID NO:118:

## (i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 588 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Met Val Leu Met Gly Arg Leu Arg Asn Ala Pro Glu Ser Leu Thr Tyr  
 1 5 10 , 15  
 40 Met Phe Cys Ala Ala Ile Arg Val Ala Pro Val Thr Thr Gln Ser Arg  
 20 25 30  
 Thr Ser Leu Arg Val Cys Thr His Val Leu Phe Pro Asp Pro Ala Leu  
 35 40 45

Pro Val Met Arg Tyr Ala Ala Asn Gly Asn Ser Arg Ser Gly Arg Pro  
 50 55 60  
 Val Gly Thr Ser Lys Ala Ala Thr Ser Arg Asn His Cys Arg Arg Gly  
 65 70 75 80  
 5 Thr Cys Val Thr Ser Ser Cys Cys Cys Glu Ser Ser Arg Met Arg Ala  
 85 90 95  
 Met Ile Gly Trp Thr Pro Cys Met Asp Val Lys Phe Lys Asn Ala Ser  
 100 105 110  
 Ser Leu Asn Arg Thr Ala Gly Leu Ala Pro Gly Cys Cys Gly Gly Gly  
 10 115 120 125  
 Pro Gly Ala Arg Thr Ser Arg Glu Pro Ser Pro Pro Asp Ala Ala Met  
 130 135 140  
 Ala Ala Gln Arg Ala Arg Ala Pro Ala Met Arg Thr Arg Gly Gly Asp  
 145 150 155 160  
 15 Ala Ala Leu Cys Ala Pro Glu Asp Gly Trp Val Lys Val His Pro Thr  
 165 170 175  
 Pro Gly Thr Met Leu Phe Arg Glu Ile Leu Leu Gly Gln Met Gly Tyr  
 180 185 190  
 Thr Glu Gly Gln Gly Val Tyr Asn Val Val Arg Ser Ser Glu Ala Ala  
 195 200 205  
 Thr Arg Gln Leu Gln Ala Ala Ile Phe His Ala Leu Leu Asn Ala Thr  
 210 215 220  
 Tyr Asp Leu Glu Glu Asp Trp Arg Arg His Val Val Arg Leu Gln Pro  
 225 230 235 240  
 25 Gln Arg Leu Val Arg Arg Tyr Arg Asn Ala Arg Glu Gly Asp Ile Ala  
 245 250 255  
 Gly Val Ala Glu Arg Val Phe Asp Thr Trp Arg Cys Thr Leu Arg Thr  
 260 265 270  
 Thr Leu Leu Asp Phe Ala His Gly Val Val Asp Cys Phe Ala Pro Gly  
 275 280 285  
 30 Gly Pro Ser Gly Pro Thr Ser Phe Pro Lys Tyr Ile Asp Trp Leu Thr  
 290 295 300  
 Cys Leu Gly Leu Val Pro Ile Leu Arg Lys Thr Arg Glu Gly Glu Ala  
 305 310 315 320  
 35 Thr Gln Arg Leu Gly Ala Phe Leu Arg Gln His Thr Leu Pro Arg Gln  
 325 330 335  
 Leu Ala Thr Val Ala Gly Ala Ala Glu Arg Ala Gly Pro Gly Leu Leu  
 340 345 350  
 Glu Leu Ala Val Ala Phe Asp Ser Thr Arg Met Ala Glu Tyr Asp Arg  
 355 360 365  
 40 Val His Ile Tyr Tyr Asn His Arg Arg Gly Glu Trp Leu Val Arg Asp  
 370 375 380  
 Pro Val Ser Gly Gln Arg Gly Glu Cys Leu Val Leu Cys Pro Pro Leu  
 349

385                      390                      395                      400  
 Trp Thr Gly Asp Arg Leu Val Phe Asp Ser Pro Val Gln Arg Leu Cys  
                          405                      410                      415  
 Pro Glu Ile Val Ala Cys His Ala Leu Arg Glu His Ala His Ile Cys  
 5                      420                      425                      430  
 Arg Leu Arg Asn Thr Ala Ser Val Lys Val Leu Leu Gly Arg Lys Ser  
                          435                      440                      445  
 Asp Ser Gly Val Ala Gly Ala Ala Arg Val Val Asn Lys Ala Leu Gly  
                          450                      455                      460  
 10 Glu Asp Asp Glu Thr Lys Ala Gly Ser Ala Ala Ser Arg Leu Val Arg  
                          465                      470                      475                      480  
 Leu Ile Ile Asn Met Lys Gly Met Arg His Val Gly Asp Ile Asn Asp  
                          485                      490                      495  
 Thr Val Arg Ala Tyr Leu Asp Glu Ala Gly Gly His Leu Ile Asp Thr  
 15                      500                      505                      510  
 Pro Ala Val Asp His Thr Leu Pro Gly Phe Gly Lys Gly Gly Thr Gly  
                          515                      520                      525  
 Arg Gly Ser Ala Ala Gln Asp Pro Gly Ala Arg Pro Gln Gln Leu Arg  
                          530                      535                      540  
 20 Gln Ala Phe Gln Thr Ala Val Val Asn Asn Ile Asn Gly Met Leu Glu  
                          545                      550                      555                      560  
 Gly Tyr Ile Asn Asn Leu Phe Gly Thr Ile Glu Arg Leu Arg Glu Thr  
                          565                      570                      575  
 Asn Ala Gly Leu Ala Thr Gln Leu Gln Ala Arg Val  
 25                      580                      585

## (2) INFORMATION FOR SEQ ID NO:119:

## (i) SEQUENCE CHARACTERISTICS:

- 30                      (A) LENGTH: 21035 base pairs  
                          (B) TYPE: nucleic acid  
                          (C) STRANDEDNESS: single  
                          (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

40 GTCTGGCCGC CGGCCCTGGC GTACGCGCTA TATAAGCCCA TGCGGTATTG GATGAGTTCC 60  
 CGCGCGCCCC GGAATCCTC CACCGCCCAC GGGGCCAGGT CCGCGGCCGC CGCGTCGAAC 120  
 TCCGCCAGCA GGCGCCCCAG GGCGTCAAAG TTCATCTCCC AGGGCACCCCT GCGCACCACC 180  
 TCATCCCACA GCCGGGCGCA CAGGGCGGTG TGCTTGGTGA CGCGCGCGCC CAGCTCCTCC 240  
 ACGGCCTCCG CGCGCTCGGC GCCCTTGGCG CCCAGGACGC CCTGGTACCT GGCGGAAAGG 300  
 CGCTCGTAGG CCGGCTGGGC CCGCAGCCCC GACACCGTGT TGGTGGTGTC CTGCAGGGCG 360

	CGCAGCTGCT	CGTGCATGGC	GCGGAACCCC	TCGGGGGACT	TCCAGGCGCC	CCCCCGGACG	420
	CGGCCAAAGC	GACCCCAGAC	CTCGTCCCAC	TCCGCCTCGG	CCTCCTCCAG	GGACCTCCGC	480
	AGGGCGTCGA	CGCGGCGCCG	AGTATCAAAG	AGCGCCCCCA	GGCGGCGGGC	GTGCCGCGCC	540
	AGGGGGCCCG	GGCCGTCGCC	GCGGGCGGCG	CTTAGCGGGT	GCGTCTCGAA	GGTGCGCTGG	600
5	GCGTGCTCTA	GCCAGATAAC	CGCGGGCACG	TCGAGCTCGC	GCGTTTCTC	GGTCTGATCC	660
	AACAGAACCT	CGACCTGGTC	GGCGATCTCC	GCCACCGAGC	GCGCCTGGTC	GAGCGTCTTG	720
	GCCACGGTCG	CCGGGACGGC	GACCACCTTC	AGCATGGTCT	TGAGGTTGGC	CAGGCCCTCG	780
	GCCTCGATCT	GGGCCC GGCG	CTCGCGCGCG	GCCAGCGCCT	CCCGCAGGCC	CGCCATGACC	840
	CGCTCGGTGG	CCTCCGCGCG	CTGCTGTTTG	GCGCGCACCA	CTGCGTCCTT	GGTCTCGGCC	900
10	GTGTCTTGCC	GGGTCACGAA	GGCGACATAC	TCGGCGTACG	CCGTGTTCTT	CACGGGGCTC	960
	TGGTCCACGC	GCTCCAACGC	CGCCGCGCAC	GCGACCAGCG	CGTCCTCGCT	GGGACACGGC	1020
	AGGGTGACCC	CGGTCCGGAC	CAGCTCCGCG	GTGGCCTCCG	GGTCATTCCG	GGCCGCGGAT	1080
	ATCTGCTCCG	CGGCGGCCGC	CAGGTCCAGG	GGCACGCCGC	CGAGCGCCCG	GTGCACGTCG	1140
	GCCCGGATGG	CGTCCAGGCG	ATCGCGGAGC	TCCACGTAGT	CGGCGTAGCC	ATGTTGGAAG	1200
15	AACGGCACGT	ACCGGCGCAG	GCCGGGCACG	CTCGTCATGT	CGTCCGCCAG	GCGCCCCACG	1260
	GCCTCGTGGT	AGTCGATAAA	CCCGTCGCCC	GCCTGGGCCA	TTTCCAGGAG	CCCCTCCGCG	1320
	ATGCGCAGCA	GCCGCGCCAG	GGGCTCGGCG	TCGACCCGAA	ACATGTCGGC	GTAGGTTTCG	1380
	GCGGCGGCGT	GGAACGCCGC	GCTCCAGCCG	AGGCGGTGGA	TGGCGGCGAG	CGGGGGGAGC	1440
	ATGGGGTGGC	GCTGGTTCTC	GGGGGTGTAG	GGGTTAAACG	CGAAGGCCGT	ATCCAGGGCG	1500
20	AGGGTGACCG	CCTCGGCGTT	GGCCGCGAGC	GCCTGCTCGG	CGCGCTTGCG	GAAGTCCCGG	1560
	GGGTGTAGC	CGTGCGTGCC	CGCCAGCGCC	TGCAGGCGGC	GCAGCTCGAC	CACGTCAAC	1620
	TCGGCGCGGT	TCTCGACGCG	GTCCAGCGCC	GCCTCGACGC	CGGCGGCCCA	GCGCTCGCTG	1680
	CTGCCCCGGG	CGCGCTGGGC	CGCCATCTTC	GCCGTCAGGT	CGGCGACGGC	GGCCTCAAGT	1740
	TCGTCGGCGC	GGCGTCGCGT	GGCGCCGATG	ACCTTGCCCA	GCTCCTGCAG	GGCGCGCCCG	1800
25	CTGGGGGAAT	GGTCCCCGGC	CGTCCCTTCG	GCGTGCAGCA	GGCCCCCGAA	CCCAGCCTCG	1860
	TGCCCCGCGA	GGCTTTCCCG	AGCAGCGGTC	GTGCGCGGGG	CCGCGGCATC	GATGAGGGCG	1920
	GCATGGTCCC	CCTCCGGCTG	GGCGCAGGCC	CGGCGCGCCT	GGACTACCAG	GTGCGCGGCC	1980
	GCCGACCCCA	GGGTCTGTAG	CTCGTCGATG	GCCCCCGCG	CCTCCAGGGC	CAGCCGAGTC	2040
	GCCTTTACAT	ACCCCGCGGC	GCTATCGGCC	AGCGCCGCGA	GGAAGGACAG	GGGCGAGGCC	2100
30	GGGTGCGGGG	CGGCCGCGCC	CAGGGCCGAC	ACCGCGTCCG	CCAGGGCGCC	ATGCGCCCCG	2160
	ACGGCCGCGT	CCACCGTCGC	CGCGGGACTT	GCCGTGCGCA	CGGCGGCGCT	CCCGGCGTTG	2220
	ATGGCGTTTG	ACACGGCTTT	GGCGATTGTG	GGGGCGTGAT	CGGAAAAGAA	CTGCACGAGG	2280
	ACCGGCGTCT	CGGGGGCGTC	GGCGAACAGG	GTCTTCAGCA	CCACCACGAA	GGCGGGATGC	2340
	AGGCCGGCCA	GAGCCGTGCG	GGTATCCGGG	GTGCGGTGTT	CCAGGGCCTC	CCGGTACTGC	2400
35	CCCAGCAGCC	CCCACAGGTC	CGCCCGCAGC	GCCGCCGTGA	CTTCCGGGGG	GGGGCCCCCG	2460
	ACGGCATCGG	CCAGGTCGGT	CCACCCCGCG	GCGAGGGAGG	CCCGCAGGGT	CGCCAGCACG	2520
	GCCGGACACG	CCTTTAGCCC	CACAAAGTCC	GGGAGGGGCC	GCAGGACCCC	TTGGAGTTTG	2580
	TGCAGGAACT	TCTCCCGGGC	GTCGTGGGCC	ACCTTGCGCG	GCTCCCGCGC	GTGTTGAGC	2640
	ATCGCCTCCA	GGGCGTGGGC	GCGTCCCGA	AGCCGGGAGC	GCGCCTCCGG	AGCGAGCTCC	2700
40	GCCGTCATCT	TGGCCGCTC	CATGGCCCTC	GCCTGCCGCA	GCGCGTCTTC	GGCCATGCGC	2760
	GTGGCCTCGG	GGGACAGCCC	GCCCCGTCG	ACGTACGGCG	CGGGGCCGGT	CGCCGGGACG	2820
	AAGGCCGCGT	CGCTGTCCAG	CTGCTGCGCG	AGCGCCGCGT	CGAGGGCGTC	GAAGCGCTGC	2880
	AGTTCGGCCA	GCCCCGAGCT	GCGCCGCGCC	TGCTGGTTCG	TGATGCCGTG	GATGCTGCGC	2940

	GCCAGCTCTT	CCAGGGGCTT	GCGTTCGATG	AGCCCCTGGG	TCGCGGCGTC	GGTCAGGACC	3000
	GAGAGCCAGG	CCGCCAGGTC	CTCGGGGGCA	TCTAGGGTCT	GGCCCCGCTG	GAGCAGGTCC	3060
	CGCAGCAGGA	TGGCCTGGGG	GCTGGTGGCG	AGGGGGGGCG	GGGGGGGGAG	CGCGGCGCGC	3120
	TGAGCGACGT	CCCGCGTGTG	TTGGTCAAAG	GCCGGTAGCG	ATTCCAGCAA	CTGGACCATG	3180
5	GGCACGACCG	CGGCCGAGGC	CACGTGAAAC	CGACAGTCGT	GGCTGTCGCT	GGCCTGCAGG	3240
	GCCTTCGCGC	TGTATACGGC	TCCCCGGTGG	AAGTACTCCT	TGATCGCGCT	CTCGATCGCC	3300
	CGGCGGGCCT	GGATCCGCAC	GTCTTCCAGC	CGCGCCTGGA	TGGCCTCGGG	GCCCAGGGCG	3360
	GGCGGGCAGC	GGGCCCTGCC	GCCGGCGCCC	GGGGCGGGCG	GCACGGGCAT	CACGGTCAGG	3420
	GGCCCCGGCG	GCTGCGAGAC	CGAGTCGACC	CCGCGGGCGA	GGGCGTCTAA	GGCCTCGCGC	3480
10	ATCTCGCGGG	CCTCCGCCTC	GACCCGCATC	TCTTCGCCCC	GGGCAAACCTG	GGCCAGCGCC	3540
	TGGATCCGAT	GGAGAAGCGG	CTCCGGGTGC	GTCGGGGTGG	CGGGGGCGAA	CAGGGTGTTC	3600
	GGGTGGGCGC	GCGAGCGCTC	CAGGAGCCAC	TCTCCGAGGC	GTGCGTACAG	ATTGGCCGGC	3660
	GGGGCGGCGC	GCAGCTGCAG	ATCCAGGTCC	GCGAGGTCCC	CGTAAAAGGC	GTCCGTCTCC	3720
	CGAATAACGT	CCCTGGCGAC	CAGGACCAGC	TTAGCGAGGG	CCAGGCGCCC	GATCTGCGAA	3780
15	TTTTCGTCCA	GCACGTGCTG	GATGAGGGGC	CGGTGGGCGG	CCACGTCCGC	CAGGCTCATG	3840
	CGCGTGGACG	CCAGGAAGTC	CCCGACGGCC	GTTTTGCGGG	GCAGCATGCG	CAGGGTGAAG	3900
	TCCAGCAGGG	CCGCGGCCGG	GCCGGCCACC	CCGGCCTGCG	TATGCGTGCG	GGCCCCGTTC	3960
	TCGATCAAAA	AGGCGAGGAC	GCGCTCAAAG	AAGAAGATGA	CGCAGAGCTC	CAACAGCCCC	4020
	GGGTGCGCCG	GGTACGGCGA	CCGCAGGGCG	TTGATGGTGA	GCTGCGAACA	CGCGGCCACC	4080
20	TCGCGGGCCA	GGGCGGCATC	GCGCGCCGCG	AGCCGGACCG	CCGTGGCGGC	CACATTGGGG	4140
	TGGACCTCGA	ACAGCTGCGC	CAGGTGCGCG	CCGGGGGGGT	CCGGGGGGCG	GCGGGCCCCC	4200
	AGCGTCTCGA	GCACGGACGG	CGACGACGGG	CTCGCGGGCC	CGTCATCGCC	GCCTCCCTGC	4260
	CCGGA CTGCG	GGGGGTATC	CGGTGCGGGA	GGGACCGTGG	CGGCTATGGG	CGTCGGGGAG	4320
	GAGGCGGGGA	CCTCGGCGGC	GACGGGGGCC	TTCTTCTTGG	GCGCGGACTT	CTTCTTGGCC	4380
25	TTGGCGGGCG	GGGCCCTTGG	GGCGGGCCTC	TCGCCCCGAGG	TCAGATCCTC	CACGCTGGAC	4440
	GGTGGGGTCC	AGGTGGGCCG	GCGGCGCTTG	GGCAAGCCGG	TAGAATAGCG	CGCCCGGTGG	4500
	CGACCCACCG	GCACTGCCCC	CACCTCCAGG	ACCCGCAGGT	CCTCGGGCTC	TTGCGCCGCG	4560
	TCCCCGGCGG	GTGTCTGCGG	GGGCGGGGCG	GCGTGCGGTG	GACCCGAGGC	CGCGGCGTCC	4620
	GGGGCCGAGG	GCTTTGCGGG	CGGGGTCCCC	TCCAGGGCTG	CTGCCCACAC	ATCATCGGGG	4680
30	GGGCGGTTTG	GGTGCCCCGC	CTGCGGTGTG	TTGGGTGGGC	CCGAGGCCCC	CCGGGGGGCC	4740
	TCGGGGGGCC	GGTCGGCCCG	AGGGGTCTGG	ACGTGGGTGG	GCGCGGGGAG	CGCGGGGACG	4800
	ACCGGGCCCG	AGCCTTCTCC	GTCCCCCCTG	GGGACCACAC	CGACAAAGAG	CGCCCTTAGC	4860
	CCCCCGATCT	CGCCCCGCAG	GGGGTGGGTG	ATGGCCACGC	GCCGCTCGAC	GAACGGTTCG	4920
	TCCTGCAGGT	AAGTCTCGCT	GGCCCCGTAG	AGGTGCAGGG	CCGCGGCGGT	CAGGTCCGCC	4980
35	GGCGCCACGG	CCCCCGGGCC	GGAGGGCACA	AAAAACACCA	TGGCGCCCGC	CCACCGCACC	5040
	TTGGGGCGGT	CGTGGGCGTA	ATACGTCAGG	TACGGGTACA	CGTCGCCCCG	CCGCACCTTG	5100
	GCGATAAACG	CGGGCGTTCC	CGCGGGCAGG	CCGTGCGGGT	CAAACAGATA	GGCCGTGTGCG	5160
	CCGTCCCGGT	AGAGCCCCAT	GCCCAGGGGG	CCGATGGTCA	GGAGCGTGTA	GGACAGCGGC	5220
	CGCATGGCCC	AGGGGCGGGC	GAAGAACGTG	TGCGCGGGGC	ATTGCGTCTC	CAGCAGCCCC	5280
40	GCCGTGGGCT	CCCCGAAGAA	GCCCACCTCG	CCGTACACCC	GCGAAAACAC	GCAACGCAGG	5340
	CCGCCGCGCG	CCGCCGGGTA	CTCCAGGAAG	TTGGGGAGCT	CGATAATGGA	ACACATGCGC	5400
	GGCGGCCCGG	AGCCCGCGGC	CGCGCGCGTC	CACTCGCCCC	CCTCCACCAG	ACATCCCTCA	5460
	ATGGCCTCCG	CGGACAGCAC	GTCGCGGGGC	CCCACGTCGA	AAAGAAGACT	GAGAAACGAC	5520

	AGGGACGAGC	GCATGCACGA	TACCGACCCC	CCCGGCTCCA	GATCGGTCGC	GAAGTGGTTC	5580
	CGAACACCGG	TGACCACGAT	ATCGCGATCC	CCCTGGCGCT	TCATCGTGGG	GTGAGGTAGC	5640
	GCGGCCGGAA	TCATGTGTGC	CGCGCCCGCC	ACGAGCGGGG	CCTGTTTATG	GGCCGGGCGT	5700
	CCCAGTGAGT	ACTGTTGTTT	CCGCCGCCCG	AACCCCCCCC	GCCCATCAAC	CGCCTGTTTCG	5760
5	TCCCCCTAAC	CACACACCCG	GTATCGCGTG	TGTGGTTTCC	CGGGAAGCCA	CATCCCACCC	5820
	CATGAAGTTT	TGCCCTTTTT	TTCCGTCCCG	CACTACGCCA	CCTTTCACC	CCCCCCCCC	5880
	AAAAAAAAAA	AAACAACAAC	CAACTCCCAG	ATGGATGGGT	GCGATAATAA	AGCTTTATTA	5940
	TTGTTTAACC	AAAGGCGAGT	CCTACGGGTG	TACCGGTGGT	GTCTCCTGCG	GCGTCATCTC	6000
	GTCGTCCTCC	ACGGGGGTGT	TGGGCCAAGG	GACCGTCTCG	CGGCCCGCCG	GGCGCGTCGA	6060
10	CGGCGCGCGG	GCCTGCGTGT	CCTGTGGGCC	GGGTGTCTGT	GGTTCGGGGG	TGCTACCGCC	6120
	GGCATCTTGG	GCCTCCAGGT	CCCCGGGGGC	CTCCGGGCCG	GCGGAAGGCC	GAAACGCCGA	6180
	GGCGCGAAAC	ACGCCGTCCG	TGACCTGCAG	GAGCTCGTTT	ATTAATAGCC	AGTCCATGCT	6240
	CAGCGTAGCG	GCCAGCCCCC	GGGGAGACAG	GTCCACGGAG	TCCGGAACCA	CCGTCCGGCTG	6300
	ACCCAGGGGC	CCCAGGCTGT	AGTCCCCCCA	GGCCCCCAGG	TCATGACGGT	TCGTGAGCAC	6360
15	GACGAGGTCT	GCGGCCGGGC	TGGGGGGCGC	GTCTCGGTC	GCGTGGGCCA	TCACCTCCTG	6420
	AATGGCTGCG	GTGCGCTGAT	CGGCCGAGCT	GGCGAAGGGC	GCCACGACCA	GCGCGCGCTC	6480
	CGTCTGCAGG	CCCTTCCACG	TGTCGTGGAG	TTCCTGAACG	AACTCGGCCA	CCCCTCGGG	6540
	GCCCCGTGGC	GCGCGCGCGG	CCTGATAGCC	GGCCGAGAGG	GCCCGCCAGC	GCGCCAGGAA	6600
	CTGACTCATG	TAACAGAACC	CGGGGACCTG	GTCCCCCGAC	ATCAACTTTG	ACGCCCTGGC	6660
20	GTGGATGCCC	GACACGATGG	CCAGGAACCC	GTGGATTTC	CGCCGCACGA	CGGCCAGCAC	6720
	GTTACCCTCG	TGCGAGACCT	GGGCCGCCAG	CTCGTCGCAT	ACCCCGAGGT	GCGCCGTCGT	6780
	CTCGGTGACG	ACGGACCGCA	GCCCCGCGAG	GGACGCGACC	AGCGCGCGCT	TGGCGTCGTG	6840
	ATACATGCCG	CAGTACTGGC	TCACCGCGTC	GCCCCATGGC	TCGGGGCGCC	AGGGCCCCAG	6900
	GCGCTCGTGG	GCGTCTGCGA	CCACGGCGTA	CAGGCGGTGC	CCGTCGCTCT	CGAACCGSCA	6960
25	CTCAAAGAAG	GCGGCGAGCG	TGCGCATGTG	CAGCCGACG	AGCACGATCG	CGTCCTCCAG	7020
	CTGGCGGACC	AGGGGGTCGG	CGCGCTCGGC	GAGTCTCTGC	AGCACCCCCC	GGGCCGCCAG	7080
	GGCGTACATG	CTGATCAGCA	GCAGGCTGCT	GCCCACCTCG	GGAGGCTGGG	GGGGAGGCAG	7140
	CTGGACCGCG	GGCCGCAGCT	GCTCGACGGC	CCCCCTGGCG	ATCACGTACA	GCTCGCGCAG	7200
	CAGTGTCTCG	ATGTTGTGCG	CCATCTGCAT	CGTGGGCCCC	ACGCCGGCCC	GGGTGGCCGG	7260
30	TTTCAGGAGG	GTGATCAGCG	CGCCCAATTT	TGTGCGGTGC	CCCTCGACGG	TGGGGAGATA	7320
	GCCCAGGCCG	AAGTCGCGCG	CCCAGGCCAG	CACCCGACGG	GCAAACCTCGA	TGGGGCGGGG	7380
	CAGGTAGGCA	GCGTTGCACG	TGGCCCTCAG	CGCGTCCCCG	ACCACCAGGG	CCAGCACGTA	7440
	AGGGACGAAC	CCCGGGTCGG	CGAGGACGTT	GGGGTGGATG	CCCTCCAGGG	CCGGGAAGCG	7500
	GATCTTGGTG	GCCGCGGCCA	GGTGAACCGA	GGGGCGGTGG	CTAGGCGGCC	CGACGGGGAG	7560
35	CAGCGCGGAC	AGCGGCGTGG	CCGGGGTGGT	GGGGGTCAGG	TCCCAGTGGG	TCTGGCCGTA	7620
	CACGTGAGC	CAGATGAGCG	CCGTCTCGCG	CAGGAGGCTG	GGCTGGCCGG	CGCTGAAGCG	7680
	GCGCTCGGCC	GTCTCAAAC	CCCCACGAG	CGTGCGCCGC	AGGCTCGCCA	GGTGTTCCTG	7740
	CGGCACGGCC	GGGCCCATGA	TGCGCGCCAG	CGTCTGGCTG	AGGACGCCGC	CCGACAGGCC	7800
	GACCGCCTCA	CAGAGCCGCC	CGTGCGTGTG	CTCGCTGGCG	CCCTGGATCC	GCCGGAACGT	7860
40	TTTCACGTAG	CCGGCGTAGT	GCCCGTACTC	CCGCGCGAGC	CCGAACACGT	TCGCCCCCGC	7920
	AAGGGCAATG	CACCCAAAGA	GCTGTGGAT	CTCGCTGAGC	CCGTGGCCGG	GGGGCGTCCG	7980
	CGCGGGCACC	CCCGCCACCA	AAAACCCCTC	CAGGGCCGAT	ATGTACTGGG	TGCAGTGCCG	8040
	GGGCGTGAAC	CCCGCGTCGG	TAAGCGTGT	GATCACCACG	GAGGGCGAGT	TGCTGTCTCTG	8100



	GACCAAAGCC	CACGTCTGCT	GCAGCAGCGC	GAGGAGCCGT	TGCTGGGCCC	CGGCGGAGGG	8160
	CGGCTCCCCT	AGCTGCAGCA	GGCCGGTGAC	GGCCGGACGG	AAGATGGCCA	GCGCCGACGC	8220
	ACTCAGAAAC	GGCACGTCGG	GGTCGAAGAC	GGCCGCGTCC	GTCCGCACGC	GCGCCATCAG	8280
	CGTCCCCGGG	GGCGCGCACG	CCGACCGCGG	GCTGACGCGG	CTTAGGGCGG	TCGACACGCG	8340
5	CACCTCCTCG	CGACTGCGAA	CCATTTTGGT	GGCCTCGAGG	GGCGGGATCA	TGATAGCCGG	8400
	GTCGATCTCC	CGCACCGTGT	GCTGAAACTG	GGCCAGCAGC	GGCGGCGGGA	CCACCGCGCC	8460
	CCGATCGGGG	GTCGTCAGGT	ACTCGTCCAC	CAGCGCCAGC	GTAACAGGG	CCC GCGTGAG	8520
	GGGGGTCAGG	GCGGCGTCGT	CGATGCGCTG	TAGGTGCGCC	GAGAACAGCG	TCACCCAATT	8580
	GCTGACCAGG	GCCAAGAACC	GGAGACCCTC	TTGCACGATC	GGGGACGCGA	AGAGCAGGCT	8640
10	GTACGCCGGG	GTGGTCAGGT	TGGCGCCGGG	TTGCCCCAGG	GGAACCGGGG	ACATCTTAAG	8700
	CGACATCTCC	CCGAGGGCCT	CCAGGGAGGT	CCGCGGGTTC	ATGGCCAGGC	AGCTCTGGGT	8760
	GACGGTCCGC	CAGCGGTCTGA	TCCACTCCAC	GGCACACTGG	CGGACGCGCA	CCGGCCCCAG	8820
	GGCCGCGGTG	GTGCGCAGCC	CGGCGGCTC	CAGCGCGTGG	GTCGTGTCGG	AGCCGGTGAT	8880
	CGCCAGGACC	GTGTCTTGA	TGACGTCCAT	CTCCCGGAAG	GCCGCTTCGG	GGGTCTCGGG	8940
15	GAGCGCCACC	GCCATGCGGT	GCACCAGCAG	CCCGGGGAGG	TTCTCGGCCA	AGAGCGCCGT	9000
	CTCCGGAAGC	CCGTGGGCCC	GGTGCAAGGC	GCACAGTTGC	TCCAGGAGCG	GGTGCCAGCA	9060
	CGCCCGCGCC	TCCGCCGGGC	CGACCGCCGC	GCCCCACAAC	AGAAACGCCG	CCGTGGCGGC	9120
	GCGCAGTTTG	GCCGCGGACA	GAAACGCCGG	CTCGTCCGCG	CTGCCCGCCG	GCTCGCTCGA	9180
	GGGGGAGGGC	GGCCGGCGGA	GGTTGGTCAG	GCTCCCCAAC	AGGACCTGCA	ACGGTCCGTT	9240
20	TGGGGGTGGA	GCGGACGGGG	GGTTCATGCC	GGCGGGCGCC	GGGACCTGGA	GCGCGCTGTC	9300
	CGACATGGCG	ACCGGCGTGC	GCGCTCGGCG	ACGCGGCGCG	GAGACGCGGG	GCCCAAACGG	9360
	GAATGACTGC	CGCCGCCCTA	TACGGAGGGG	CTAAGTATCG	CCCGGGGACC	CTTCGAAACC	9420
	CCGGGCGTGT	CGCAAGTACG	CCGCGAAGGC	GCGGCGTGTT	ATACGGCGCG	TTATGTCCCG	9480
	GCATTCCGTT	CGTGGGTTCG	GGCCCCGGTG	CTGTGCGGTG	GGAGTGTGTG	TGTGTGGGGG	9540
25	GGGGGCGGCG	CGACGGCGGC	CCGGACCAAG	TGTATCGCGG	CCGTTCCGTG	GGGCGGCCCA	9600
	ACAGGCCCTT	TAAACATTTG	CGTATGCACC	GGCCCAGCCA	GTCGGACACC	GGAACCCACC	9660
	AGAGGCGGAA	GCCGCTTCG	CCCGTGAGGG	TGCGTGTGTT	TTCTGGTGGC	GTGTTTTTCC	9720
	TTTCCGCCCT	CCTCCCTCCC	CACCTCCACC	ACCCCCCCCC	CACAACTCGC	CCGTTGGCGA	9780
	TCGGCGGGAA	AACCATGAAA	ACCAAGCCAC	TCCCGACAGC	CCCGATGGCG	TGGGCCGAGA	9840
30	GTGCCGTGGA	AACCACCACC	AGCCCGCGCG	AGCTCGCGGG	CCACGCCCCG	CTCCGGCGCG	9900
	TCCTGCGCCC	GCCCATCGCT	CGCCGCGACG	GCCCGGTGCT	TTTGGGGGAC	AGGGCCCCCA	9960
	GGAGGACGGC	CAGTACGATG	TGGCTGCTGG	GGATCGACCC	CGCGGAGTCG	TCTCCGGGAA	10020
	CGCGCGCTAC	CCGAGACGAT	ACCGAGCAGG	CCGTGGACAA	GATCCTCAGG	GGAGCCCCGG	10080
	GCGCGGGAGG	GCTGACCGTC	CCCGGCGCCC	CCCGCTATCA	CCTGACCCGC	CAGGTAACCC	10140
35	TGACGGATCT	CTGCCAACCA	AACGCGGAGC	GGGCGGGGCG	GCTCCTTTTG	GCCCTGCGGC	10200
	ACCCACACGA	CCTCCCCCAC	CTGGCCCCGC	ATCGGGCTCC	GCCCGGCCGG	CAGACCGAGC	10260
	GACTGGCCGA	GGCTGGGGC	CAGCTCCTGG	AGGCCTCCGC	CCTGGGGTCC	GGGCGGGCCG	10320
	AGAGCGGCTG	CGCGCGCGCG	GGCCTTGTGT	CGTTTAACTT	TCTGGTGGCC	GCGTGCGCCG	10380
	CCGCTTACGA	TGCGCGCGAC	GCCGCCGAGG	CGGTCCGGGC	CCACATCACG	ACCAACTACG	10440
40	GCGGGACGCG	GGCCGGGGCG	CGGCTGGACC	GGTTTTCCGA	ATGCCTGCGC	GCCATGGTCC	10500
	ACACGCACGT	GTTTCCCCAC	GAGGTCATGC	GGTTTTTCGG	GGGGCTAGTG	TCGTGGGTCA	10560
	CACAGGACGA	GCTGGCTAGC	GTCACCGCCG	TCTGCAGCGG	ACCCAGGAG	GCCACACACA	10620
	CCGGCCACCC	GGGAGGCCCC	TGTTTCGGCCG	TTACCATCCC	GGCCTGCGCC	TTCGTGGACC	10680

	TGGACGCCGA	GCTGTGCCTG	GGGGGCCCTG	GGGCGGCGTT	CCTGTACTTG	GTCTTCACCT	10740
	ACCGACAGTG	CCGGGACCAG	GAGCTCTGTT	GCGTGTACGT	GGTCAAGAGC	CAGCTCCCCC	10800
	CGCGCGGACT	GGAGGCGGCC	CTCGAGCGGC	TGTTGCGGCG	CCTCCGGATA	ACCAACACGA	10860
	TTACAGGGGC	CGAGGACATG	ACGCCCCCTC	CCCCGAACCG	AAACGTTGAC	TTTCCGCTCG	10920
5	CCGTCTTGGC	CGCGAGCTCG	CAATCCCCGC	GGTGCTCGGC	GAGCCAAGTC	ACGAACCCCC	10980
	AGTTTGTGCA	CAGGCTGTAC	CGCTGGCAGC	CGGATCTGCG	GGGGCGCCCT	ACCGCACGCA	11040
	CCTGCACATA	CGCCGCCTTC	GCAGAGCTGG	GTGTATGCC	AGACAACAGC	CCCCGCTGTC	11100
	TGCACCGCAC	CGAGCGGTTT	GGGGCGGTGC	GCGTTCCGGT	TGTCATCCTG	GAGGGCGTGG	11160
	TGTGGCGCCC	CGGCGGGTGG	CGGGCCTGCG	CGTGATCGTC	TATTGACGAC	GGCCGCCCAA	11220
10	CCCGAGCGAC	CTTCCCCCTC	CACTTTCCCC	CCCCCCCCCTC	CTACACACCA	ACTCCGCCCT	11280
	CGCCGCTCTG	GCCGTGCGCG	GCCCCGTGCG	TCCGTCTCAA	TAAAGCCAGG	TTAAATCCGT	11340
	GACGTGGTGT	GTTTGGCGTG	TGTCTCTGAA	ATGGCGGAAA	CCGACATGCA	AATGGGATTC	11400
	ATGGACACGT	TACACCCCCC	TGACTCAGGA	GATAGGCATA	TCCTCCTTAG	ATTGACTCAG	11460
	CACACGATCG	CACCCACACC	CTGTGTGCCG	GGGATAAAAG	CCAACGCGGG	CGGTCTGGGT	11520
15	TACCACAACA	GGTGGGTGCT	TCGGGGACTT	GACGGTCGCC	ACTCTCCTGC	GAGCCCTCAC	11580
	GTCTTCGCCC	ACCGATTCTT	GTTGCGTTCC	TGTCGGCCGG	TGCTGTCTCG	TCGACAGATT	11640
	GTTGGCGACT	GCCCCGGTGA	TTCGTGCGCC	GGTGCGTCCT	TTGCGTCGTA	CCGCCCACCC	11700
	CGCCTCCAC	GGGCCCCCG	CTGTTTCCGT	TCATCGCGTC	CGAGCCACCG	TCACCTTGGT	11760
	TCCAATGGCC	AACCGCCCTG	CCGCATCCGC	CCTCGCCGGA	GCGCGGTCTC	CGTCCGAACG	11820
20	ACAGGAACCC	CGGGAGCCCC	AGGTGCCCCC	CCCTGGCGGC	GACCACGTGT	TTTGACAGAA	11880
	AGTCAGCGGC	GTGATGGTGC	TTTCCAGCGA	TCCCCCGGCG	CCCGCGGCCT	ACCGCATTAG	11940
	CGACAGCAGC	TTTGTTCAT	GCGGCTCCAA	CTGCAGTATG	ATAATCGACG	GAGACGTGGC	12000
	GCGCGGTGAT	TTGCGTGACC	TCGAGGGCGC	TACGTCCACC	GGCGCCTTCG	TCGCGATCTC	12060
	AAACGTCGCA	GCCGGCGGGG	ATGGCCGAAC	CGCCGTGCTG	GCGCTCGGCG	GAACCTCGGG	12120
25	CCCGTCCGCG	ACTACATCCG	TGGGGACCCA	GACGTCCGGG	GAGTTCCTCC	ACGGGAACCC	12180
	AAGGACCCCC	GAACCCCAAG	GACCCACAGC	TGTCCCCCGG	CCCCCTCCTC	CCCCCTTTCC	12240
	ATGGGGCCAC	GAGTGCTGCG	CCCCTGCGCA	TGCCAGGGGC	GGCGCCGAGA	AGGACGTGCG	12300
	GGCCGCGGAG	TCATGGTCAG	ACGGCCCGTC	TCCGACTCC	GAAACGGAGG	ACTCGGACTC	12360
	CTCGGACGAG	GATACGGGCT	CGGGTTCCGA	GACGCTGTCT	CGATCCTCTT	CGATCTGGGC	12420
30	CGCAGGGGCG	ACTGACGACG	ATGACAGCGA	CTCCGACTCG	CGGTGCGACG	ACTCCGTGCA	12480
	GCCCCACGTT	GTCGTTGCTC	GCAGATGGAG	CGACGGCCCT	GCCCCCGTGG	CCTTTCCCAA	12540
	GCCCCGCGCG	CCCGGCGACT	CCCCCGGAAA	CCCCGGCCTG	GGCGCCGGCA	CCGGGCGGGG	12600
	CTCCGCGACG	GACCCGCGCG	CGTCGGCCGA	CTCCGATTCC	GCGGCCACG	CCGCCGCACC	12660
	CCAGGCGGAC	GTGGCGCCGG	TTCTGGACAG	CCAGCCCACT	GTGGGAACGG	ACCCCGGCTA	12720
35	CCCAGTCCCC	CTAGAACTCA	CGCCCCGAGAA	CGCGGAGGCG	GTGGCGCGGT	TTCTGGGGGA	12780
	CGCCGTCGAC	CGCGAGCCCC	CGCTCATGCT	GGAGTACTTC	TGTCGGTGCG	CCCGCGAGGA	12840
	GAGCAAGCGC	GTGCCCCCAC	GAACCTTCGG	CAGCGCCCCC	CGCCTCACGG	AGGACGACTT	12900
	TGGGCTCCTG	AACTACGCGC	TCGCTGAGAT	GCGACGCCTG	TGCCTGGACC	TTCCCCCGGT	12960
	CCCCCCCCAAC	GCATACACGC	CCTATCATCT	GAGGGAGTAT	GCGACGCGGC	TGGTTAACGG	13020
40	GTTCAAACCC	CTGGTGCGGC	GGTCCGCCCG	CCTGTATCGC	ATCCTGGGGA	TTCTGGTTCA	13080
	CCTGCGCATC	CGTACCCGGG	AGGCCTCCTT	TGAGGAATGG	ATGCGCTCCA	AGGAGGTGGA	13140
	CCTGGACTTC	GGGCTGACGG	AAAGGCTTCG	CGAACACGAG	GCCCAGCTAA	TGATCCTGGC	13200
	CCAGGCCCTG	AACCCCTACG	ACTGTCTGAT	CCACAGCACC	CCGAACACGC	TCGTGAGCG	13260

	GGGGCTGCAG	TCGGCGCTGA	AGTACGAAGA	GTTTTACCTC	AAGCGCTTCG	GCGGGCACTA	13320
	CATGGAGTCC	GTCTTCCAGA	TGTACACCCG	CATCGCCGGG	TTCCTGGCGT	GCCGGGCGAC	13380
	CCGCGGCATG	CGCCACATCG	CCCTGGGSCG	ACAGGGGTCG	TGGTGGGAAA	TGTTCAAGTT	13440
	CTTTTTCCAC	CGCCTCTACG	ACCACCAGAT	CGTGCCGTCC	ACCCCCGCCA	TGCTGAACCT	13500
5	CGGAACCCGC	AACTACTACA	CGTCCAGCTG	CTACCTGGTA	AACCCCAGG	CCACCACTAA	13560
	CCAGGCCACC	CTCCGGGCCA	TCACCGGCAA	CGTGAGCGCC	ATCCTCGCCC	GCAACGGGGG	13620
	CATCGGGCTG	TGCATGCAGG	CGTTCAACGA	CGCCAGCCCC	GGCACCGCCA	GCATCATGCC	13680
	GGCCCTGAAG	GTCTTGGA	CCCTGGTGGC	GGCGCACAA	AAACAGAGCA	CGCGCCCCAC	13740
	CGGGGCGTGC	GTGTACCTGG	AACCCTGGCA	CAGCGACGTT	CGGGCCGTGC	TCAGAATGAA	13800
10	GGGCGTCCTC	GCCGGCGAGG	AGGCCAGCG	CTGCGACAAC	ATCTTCAGCG	CCCTCTGGAT	13860
	GCCGGACCTG	TTCTTCAAGC	GCCTGATCCG	CCACCTCGAC	GGCGAGAAAA	ACGTCACCTG	13920
	GTCCCTGTTC	GACCGGACA	CCAGCATGTC	GCTCGCCGAC	TTTCACGGCG	AGGAGTTCGA	13980
	GAAGCTGTAC	GAGCACCTCG	AGGCCATGGG	GTTCCGGCGAA	ACGATCCCCA	TCCAGGACCT	14040
	GGCGTACGCC	ATCGTGC	GCGCGGCCAC	CACCGGAAGC	CCCTTCATCA	TGTTTAAGGA	14100
15	CGCGGTAAAC	CGCCACTACA	TCTACGACAC	GCAAGGGGCG	GCCATTGCCG	GCTCCAACCT	14160
	CTGCACGGAG	ATCGTCCACC	CGTCTCCAA	ACGCTCCAGC	GGGGTCTGCA	ACCTGGGCAG	14220
	CGTGAATCTG	GCCCGATGCG	TCTCCCGCG	GACGTTTCGAT	TTTGGCATGC	TCCGCGACGC	14280
	CGTGCAGGCG	TGCGTGCTAA	TGGTTAATAT	CATGATAGAC	AGCACGCTGC	AGCCGACGCC	14340
	CCAGTGCGCC	CGCGGCCACG	ACAACCTGCG	GTCCATGGGC	ATTGGCATGC	AGGGCCTGCA	14400
20	CACGGCGTGC	CTGAAGATGG	GCCTGGATCT	GGAGTCGGCC	GAGTTCGGG	ACCTGAACAC	14460
	ACACATCGCC	GAGGTGATGC	TGCTCGCGGC	CATGAAGACC	AGTAACGCGC	TGTGCGTTTCG	14520
	CGGGGCGCGT	CCCTTCAGCC	ACTTTAAGCG	CAGCATGTAC	CGGGCCGGCC	GCTTTCAGTG	14580
	GGAGCGCTTT	TCGAACGCCA	GCCCGCGGTA	CGAGGGCGAG	TGGGAGATGC	TACGCCAGAG	14640
	CATGATGAAA	CACGGCCTGC	GCAACAGCCA	GTTCATCGCG	CTCATGCCCA	CCGCCGCCTC	14700
25	GGCCCAGATC	TCGGACGTCA	GCGAGGGCTT	TGCCCCCTG	TTCACCAACC	TGTTTCAGCAA	14760
	GGTGACCAGG	GACGGCGAGA	CGCTGCGCCC	CAACACGCTC	TTGCTGAAGG	AACTCGAGCG	14820
	CACGTTCCGC	GGGAAGCGGC	TCCTGGACGC	GATGGACGGG	CTCGAGGCCA	AGCAGTGGTC	14880
	TGTGGCCCAG	GCCCTGCCTT	GCCTGGACCC	CGCCACCCC	CTCCGGCGGT	TCAAGACGGC	14940
	CTTCGACTAC	GACCAGGAAC	TGCTGATCGA	CCTGTGTGCA	GACCGCGCCC	CCTATGTTGA	15000
30	TCACAGCCAA	TCCATGACTC	TGTATGTAC	AGAGAAGGCG	GACGGGACGC	TCCCCGCCTC	15060
	CACCCTGGTC	CGCCTTCTCG	TCCACGCATA	TAAGCGCGGC	CTGAAGACGG	GGATGTACTA	15120
	CTGCAAGGTT	CGCAAGGCGA	CCAACAGCGG	GGTGTTCGCC	GGCGACGACA	ACATCGTCTG	15180
	CACAAGCTGC	GCGCTGTAAG	CAACAGCGCT	CCGATCGGGG	TCAGGCGTCG	CTCTCGGTCC	15240
	CGCATATCGC	CATGGATCCC	GCCGTCTCCC	CCGCGAGCAC	CGACCCCTA	GATACCCACG	15300
35	CGTCGGGGGC	CGGGGCGGCC	CCGATTCGGG	TGTGCCCCAC	CCCCGAGCGG	TACTTCTACA	15360
	CCTCCAGTG	CCCCGACATC	AACCACCTTC	GCTCCCTCAG	CATCCTGAAC	CGCTGGCTGG	15420
	AGACCGAGCT	CGTGTTCGTG	GGGGACGAGG	AGGACGTCTC	CAAGCTCTCC	GAGGGCGAGC	15480
	TCGGCTTCTA	CCGCTTCTTG	TTTGCCTTCC	TGTGCGCCGC	GGACGACCTG	GTGACGGAAA	15540
	ACCTGGGCGG	CCTCTCCGGC	CTCTTCGAAC	AGAAGGACAT	TCTTCACTAC	TACGTGGAGC	15600
40	AGGAATGCAT	CGAGGTCGTC	CACTCGCGCG	TCTACAACAT	CATCCAGCTG	GTGCTCTTTC	15660
	ACAACAACGA	CCAGGCGCGC	CGCGCCTATG	TGGCCCGCAC	CATCAACCAC	CCGGCCATTC	15720
	GCGTCAAGGT	GGACTGGCTG	GAGGCGCGGG	TGCGGGAATG	CGACTCGATC	CCGGAGAAGT	15780
	TCATCCTCAT	GATCCTCATC	GAGGCGGTCT	TTTTTGCCGC	CTCGTTCGCC	GCCATCGCGT	15840

	ACCTGCGCAC	CAACAACCTC	CTGCGGGTCA	CCTGCCAGTC	GAACGACCTC	ATCAGCCGCG	15900
	ACGAGGCCGT	GCATACGACA	GCCTCGTGCT	ACATCTACAA	CAACTACCTC	GGGGGCCACG	15960
	CCAAGCCCGA	GGCGGCGCGC	GTGTACCGGC	TGTTTCGGGA	GGCGGTGGAT	ATCGAGATCG	16020
	GGTTCATCCG	ATCCCAGGCC	CCGACGGACA	GCTCTATCCT	GAGTCCGGGG	GCCCTGGCGG	16080
5	CCATCGAGAA	CTACGTGCGA	TTCAGCGCGG	ATCGCCTGCT	GGGCCTGATC	CATATGCAGC	16140
	CCCTGTATTC	CGCCCCCGCC	CCCGACGCCA	GCTTTCCTCT	CAGCCTCATG	TCCACCGACA	16200
	AACACACCAA	CTTCTTCGAG	TGCCGCAGCA	CCTCGTACGC	CGGGGCCGTC	GTCAACGATC	16260
	TGTGAGGGTC	TGGGCGCCCT	TGTAGCGATG	TCTAACCGAA	ATAAAGGGGT	CGAAACGGAC	16320
	TGTTGGGTCT	CCGGTGTGAT	TATTACGCAG	GGGAGGGGGG	TGGCGGTGG	GGAAAGGGAA	16380
10	GGAACGCCCG	AAACCAGAGA	AAAGGACCAA	AAGGGAAACG	CGTCCAACCG	ATAAATCAAG	16440
	CGCCGACCAG	AACCCCGAGA	TGCATAATAA	CAAACGATTT	TATTACTCTT	ATTATTAACA	16500
	GGTCGGGCAT	CGGAGGGGGA	TGGGGGCGCG	CGTTTCCTCC	GTCCCGGCTA	CTCGTCCCGA	16560
	AATTTAGCCA	GGACGTCTTT	GTAAACGCG	GGCGGGGGCG	CGTGGGCCCA	CAGCTGCGCC	16620
	AGAAACCGGT	CGGCGATGTC	CGGGGCGGTG	ATATGCCGAG	TCACGATGGA	GCGCGCTAAA	16680
15	TCTTCGTGCG	GGAGGTCTTG	ATAGATGGGC	AGTCTTTTTA	GAAGAGTCCA	GGGTCCCCGC	16740
	TCCTTGGGGC	TGATAAGCGA	TATGACGTAC	TTGACGTATC	TGTGCTCCAC	CAGCTCGGCG	16800
	ATGGTCATCG	GATCGGGCAG	CCAGTCCAGG	GCCTCCGGGG	CGTCGTGGAT	GACGTGGCGG	16860
	CGACGTCCGG	CGACATAGCC	GCGGTGTTCC	GCGACCCGCT	GCGCGTTGGG	GACCTGCACC	16920
	AGCTCGGGCG	GGGTGAGTAT	CTCCGAGGAG	GACGACCGGG	CGCCGTGCGG	CGGCCCACCG	16980
20	GCGACGTCCG	GGGGCTGGAG	GGGGGGGTCT	TCTTCGTAGT	CGTCTCGCC	CGCGATCTGT	17040
	TGGGCCAGAA	TTTCGGTCCA	CGAGATGCGC	GTCTCGAGGC	CGACCGGGGC	CGCGGTGAGC	17100
	GTAGGCATGC	TCTCCAGGGA	GCGCGAGTTG	GCGCGCTCCC	GCCGGGCCGC	CCGGCGGGCC	17160
	TGGGATCGGC	TCGGGGCGGT	CCAGTGACAC	TCGCGCAGCA	CGTCTCGAC	GGACGCGTAG	17220
	GTGTTATTGG	GGTGCAGGTC	TGTGTGGCAG	CGGACGAACA	GCGCCAGGAA	CTGCGGGTAA	17280
25	CTCATCTTGA	AGTACTGCAG	CAGGTGCGCG	CAGTGAATCG	TCGGAATGTA	GCCGGTGCTG	17340
	ATGTCCAACA	CGATATCGCA	GCCCATCAGC	AGGAGATCGG	TATCCGTGGT	ATGCACGTAC	17400
	GCGACCGTGT	TGGTATGATA	GAGGTTCGCG	CAGGCGTCGT	CGGCCTCCAG	CTGACCCGAG	17460
	TTGATGTAGG	CGTACCCAG	CGCCCGCAGA	ACGCGGATAC	AGAACAGGTG	AGCCAGGCGC	17520
	AGGGCCGGCT	TCGAGGGCGC	GCCCCAGGGG	GCCGCCGGGC	CTGGGCCGGC	GGCCCCGCTT	17580
30	CCCCGGTCCC	CCGGGGCGAA	GGCGTGCCCG	CGGCGGCGCA	TGTTGGAAAA	AGGCGAAACT	17640
	GGGCCTGGAG	TCGGTGATGG	GGGAAGGCGG	CGGCGAGGCG	TCTACGTAC	TGGCCTCCTC	17700
	GTCCGTGCGG	CACTGGGCCG	TCGTGCGGGC	CAGGATCGCC	TTGGCCCCGA	ACACAACCGG	17760
	CTCGGTACAC	TCGACCCCGC	GATCGGTAC	GAAGATGGGG	AACAGGGACT	TTTGGGTAAA	17820
	CACCCGTAAC	ATACTACAGA	GACAGTGTAG	CGTGATTGCC	TCGCGGTCTG	AACTTGGGTA	17880
35	GCGGCGCTGA	TATTTAACCA	CCAGGGTATA	CATGACATTC	CACAGGTCCA	CGGCGATGGG	17940
	GGTAAAGTAG	CCCTCCGGGG	CCCGGAGGCC	CCGGCGCTTC	ACCAGATGGT	GAGTCTGGGC	18000
	AAACTTCATC	ATGCCAAACA	GACCCATTC	GGCACGATTG	TAGGTGCGGA	TAGGTCTCTC	18060
	TACAGAGCTG	TATAGGTGTG	ACGGTCCGGG	ACACCAAGC	CCGCCGCCCC	TGTGTACAGT	18120
	GGCTGCGGCG	ACGACCCCGC	TCCAACAAGA	CGCTATCCCG	GGAAGGCAC	GCTCTTTATA	18180
40	ATTCTTTTTT	ATTTCCCATC	TACGTGCGGA	TTGGTGCAAC	CGCCGGCGCG	CGCCGGTGCA	18240
	GGCCGACCAT	CTCTCTCTTC	CCCCCTCCC	CCTCCCCCGA	GCCCTCAAAG	AGGGTGTGGC	18300
	CTAACTAGCG	GAAGGCGTAT	TTAACCAGAC	TAGGCGGCG	GGTCCGCCGT	AGTCTTGGC	18360
	TCGGGTAGCC	ACTGCTCTGT	GGCTCGGGTC	CCCCGGCCCC	CCTAACCCCC	ATCCGGTCCG	18420

	CGTCATCCGC	CCCCCTCCGCC	TGCGACACAA	ACGGCCGCGC	CTCCGGGCCC	GGTGACACGA	18480
	CGCGCCTCGT	CTCTGCGGAT	TGTCCCGGGA	GCGTCGCGGC	ATGGCTCATC	TTCCCGGCGG	18540
	TGCGGCCGCC	CCCCCCTTT	CGGAGGAEGC	GATCCCGTCG	CCGCGCGAGC	GGACGGAAGA	18600
	CTGGCCGCCC	TGCCAGATAG	TGCTGCAGGG	CGCCGAGCTG	AACGGGATCC	TGCAGGCCTT	18660
5	TGCGCCGCTT	CGCACGAGCC	TTTTGGACTC	GCTCCTGGTC	GTGGGCGACC	GAGGCATCCT	18720
	TGTACATAAC	GCGATTTTCG	GCGAGCAGGT	GTTTCTGCCC	CTCGACCATT	CGCAGTTCAG	18780
	TCGCTATCGA	TGGGGCGGAC	CCACCGCGGC	GTTCTGTCT	CTCGTGAGCC	AGAAGCGATC	18840
	CCTGCTGAGC	GTGTTTCGCG	CCAACCAGTA	CCCTGACCTG	CGGCGGGTGG	AGCTGACGGT	18900
	CACGGGCCAG	GCCCCGTTTC	GCACGCTGGT	GCAGCGCATA	TGGACGACCG	CGTCCGACGG	18960
10	AGAGGCCGTG	GAGCTTGCCA	GCGAGACGCT	CATGAAACGC	GAGTTGACGA	GCTTCGCGGT	19020
	ACTACTCCCC	CAGGGCGACC	CCGACGTCCA	GCTGCGCCTC	ACGAAGCCCC	AGCTCACGAA	19080
	GGTGGTGAAC	GCCGTCGGGG	ACGAGACCGC	CAAACCCACC	ACGTTCGAGC	TCGGCCCCAA	19140
	CGGCAAGTTT	TCCGTGTTTA	ACGCGCGCAC	CTGCGTCACC	TTTGCCGCCC	GCGAGGAGGG	19200
	CGCGTCGTCC	AGCACCAGCG	CCCAGGTCCA	GATTCTGACC	AGCGCGCTGA	AGAAGGCGGG	19260
15	CCAGGCGGCC	GCCAACGCCA	AGACGGTCTA	CGGGGAAAC	ACACACCGCA	CATTCTCGGT	19320
	GGTCGTCGAC	GA CTGCAGCA	TGCGGGCGGT	CCTCCGGCGG	CTCCAGGTCG	GCGGGGGGAC	19380
	CCTCAAGTTC	TTCTCACGG	CCGACGTCCC	CAGCGTGTGT	GTCACCGCCA	CCGGCCCCAA	19440
	CGCGGTGTCG	GCGGTGTTTC	TTTTAAACC	CCAGCGGGTC	TGCCTGAACT	GGCTCGGCCG	19500
	GACCCCGGT	TCCTCGACCG	GGAGCTTGGC	GTCCACGAC	TCTCGGGCCG	GCCCGACCGA	19560
20	CAGCCAGGAC	TTCTCTCCG	AGCCGGACGC	GGGCGACCGC	GGCGCCCCAG	AAGAAGAAGG	19620
	CCTCGAGGGC	CAGGCCCGGG	TCCCGCCCGC	GTTCCCGGAA	CCGCCGGGAA	CCAAGCGGAG	19680
	GCACGCCGGG	GCCGAAGTTG	TCCCGCGGGA	CGACGCCACC	AAGCGCCCGA	AGACGGGCGT	19740
	GCCCGCCGCC	CCCACGCGAG	CCGAGTCGCC	CCCCCTCTCC	GCGAGATACG	GACCCGAGGC	19800
	GGCGGAGGGT	GGTGGGGACG	GCGGCCGCTA	CGCGTGCTAC	TTTCGCGACC	TCCAGACCGG	19860
25	CGACGCGAGC	CCCAGCCCCC	TCTCCGCCCT	CCGGGGTCCC	CAAAGACCCC	CATACGGCTT	19920
	TGGGTTGCCC	TGACGCGGAC	GGGTGGTGGC	CGAACGCTTC	ACCGCGCCCC	GGCAGCGGGG	19980
	GTGCGTTGTG	TTAAAAAAT	AAATAAATGG	GGTAGTGTGT	CCCCCCCCCTC	CAACCAATAT	20040
	GGCTGTGCTG	TGTGTTCCG	GGTTGCGCCT	CCGTCTTTTC	CACCCCCCTT	CCCCCTCCTT	20100
	TTTTGTTTTG	CGTGCGCTTA	TAAGAGCGGG	CCCGGGGCCC	TTTCGAGCTT	CACCGAGAGC	20160
30	GCCGTCGGGC	CCCGGGTGCG	GGATGTGTCG	CGGGGACAGC	CCCGGGGTCG	CGGGCGGGAG	20220
	CGGCGAACAC	TGCCTCGGAG	GGGATGATGG	GGACGACGGG	CGCCCCCGCC	TCGCCTGCGT	20280
	GGGTGCCATC	GCTCGGGGGT	TCGCGCATCT	CTGGCTCCAG	GCCACCACGC	TGGGCTTCGT	20340
	GGGGTCTGTC	GTTCTGTGCG	GCGGCCCGTA	TGCGGACGCC	ATGTCGGGGG	CGTTCGTGAT	20400
	CGGGAGCACC	GGCCTGGGGT	TCCTCCGCGC	CCCCCCCCCG	TTTCGCCCCG	CGCCGACGCG	20460
35	TGTGTGCGCG	TGGCTGAGGC	TGGTCGGCGG	GGGAGCGGCC	GTGGCCCTGT	GGAGCCTCGG	20520
	GGAGGCCGGC	GCGCCTCCGG	GGGTTCGGGG	CCCGCGGACC	CAGTGCCTGG	CGCTCGGGGC	20580
	CGCCTACGCG	GCGCTGCTGG	TGCTGGCCGA	CGACGTCCAT	CCCTTTTTC	TCCTCGCCCC	20640
	GCGGCCCTG	TTTGTGCGCA	CCCTGGGGGT	TGTCGTGCGC	GGGCTGACGA	TAGGCGGCAG	20700
	TGCGCGCTAC	TGGTGGATCG	ACCCCCGCGC	CGCCGCGGCC	CTGACGGCGG	CGGTGGTGGC	20760
40	GGGCCTCGGG	ACAACCGCCG	CCGGGGACAG	CTTTTCCAAG	GCCTGTCCCC	GCCACCGCCG	20820
	CTTTTGCCTG	GTCTCCGCGG	TCGAGTCTCC	CCCGCCCCGA	TACGCCCCCG	AGGACGCCGA	20880
	GCGGCCAACA	GACCACGGAC	CCCTGTTACC	GTCGACGCAC	CACCAGCGAT	CTCCGCGGGT	20940
	CTGCGGCGAC	GGGGCCGCAC	GGCCCGAAAA	CATCTGGGTT	CCCGTGGTGA	CCTTTGCGGG	21000

CGCGCTCGCG CTGGCCGCCT GCGCCGCGCG AGGG

21035

## (2) INFORMATION FOR SEQ ID NO:120:

- 5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1850 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

15 Val Ala Gly Ala Ala His Met Ile Pro Ala Ala Leu Pro His Pro Thr  
 1 5 10 15  
 Met Lys Arg Gln Gly Asp Arg Asp Ile Val Val Thr Gly Val Arg Asn  
 20 20 25 30  
 Gln Phe Ala Thr Asp Leu Glu Pro Gly Gly Ser Val Ser Cys Met Arg  
 20 35 40 45  
 Ser Ser Leu Ser Phe Leu Ser Leu Leu Phe Asp Val Gly Pro Arg Asp  
 50 55 60  
 Val Leu Ser Ala Glu Ala Ile Glu Gly Cys Leu Val Glu Gly Gly Glu  
 65 70 75 80  
 25 Trp Thr Arg Ala Ala Gly Ser Gly Pro Pro Arg Met Cys Ser Ile  
 85 90 95  
 Ile Glu Leu Pro Asn Phe Leu Glu Tyr Pro Ala Arg Gly Leu Arg Cys  
 100 105 110  
 Val Phe Ser Arg Val Tyr Gly Glu Val Gly Phe Phe Gly Glu Pro Thr  
 115 120 125  
 30 Ala Gly Leu Leu Glu Thr Gln Cys Pro Ala His Thr Phe Phe Ala Gly  
 130 135 140  
 Pro Trp Ala Met Arg Pro Leu Ser Tyr Thr Leu Leu Thr Ile Gly Pro  
 145 150 155 160  
 35 Leu Gly Met Gly Arg Asp Gly Asp Thr Ala Tyr Leu Phe Asp Pro His  
 165 170 175  
 Gly Leu Pro Ala Gly Thr Pro Ala Phe Ile Ala Lys Val Arg Ala Gly  
 180 185 190  
 Asp Val Tyr Pro Tyr Leu Thr Tyr Tyr Ala His Asp Arg Pro Lys Val  
 195 200 205  
 40 Arg Trp Ala Gly Ala Met Val Phe Phe Val Pro Ser Gly Pro Gly Ala  
 210 215 220  
 Val Ala Pro Ala Asp Leu Thr Ala Ala Ala Leu His Leu Tyr Gly Ala

359

225                      230                      235                      240  
 Ser Glu Thr Tyr Leu Gln Asp Glu Pro Phe Val Glu Arg Arg Val Ala  
                                  245                      250                      255  
 Ile Thr His Pro Leu Arg Gly Glu Ile Gly Gly Leu Gly Ala Leu Phe  
 5                                   260                      265                      270  
 Val Gly Val Val Pro Arg Gly Asp Gly Glu Gly Ser Gly Pro Val Val  
                                  275                      280                      285  
 Pro Ala Leu Pro Ala Pro Thr His Val Gln Thr Pro Arg Ala Asp Arg  
                                  290                      295                      300  
 10 Pro Pro Glu Ala Pro Arg Gly Ala Ser Gly Pro Pro Asn Thr Pro Gln  
 305                                   310                      315                      320  
 Ala Gly His Pro Asn Arg Pro Pro Asp Asp Val Trp Ala Ala Ala Leu  
                                  325                      330                      335  
 Glu Gly Thr Pro Pro Ala Lys Pro Ser Ala Pro Asp Ala Ala Ala Ser  
 15                                   340                      345                      350  
 Gly Pro Pro His Ala Ala Pro Pro Pro Gln Thr Pro Ala Gly Asp Ala  
                                  355                      360                      365  
 Ala Glu Glu Ala Glu Asp Leu Arg Val Leu Glu Val Gly Ala Val Pro  
                                  370                      375                      380  
 20 Val Gly Arg His Arg Ala Arg Tyr Ser Thr Gly Leu Pro Lys Arg Arg  
 385                                   390                      395                      400  
 Arg Pro Thr Trp Thr Pro Pro Ser Ser Val Glu Asp Leu Thr Ser Gly  
                                  405                      410                      415  
 Glu Arg Pro Ala Pro Lys Ala Pro Pro Ala Lys Ala Lys Lys Lys Ser  
 25                                   420                      425                      430  
 Ala Pro Lys Lys Lys Ala Pro Val Ala Ala Glu Val Pro Ala Ser Ser  
                                  435                      440                      445  
 Pro Thr Pro Ile Ala Ala Thr Val Pro Pro Ala Pro Asp Thr Pro Pro  
                                  450                      455                      460  
 30 Gln Ser Gly Gln Gly Gly Gly Asp Asp Gly Pro Asp Ser Ser Pro Ser  
 465                                   470                      475                      480  
 Val Leu Glu Thr Leu Gly Ala Arg Arg Pro Pro Glu Pro Pro Gly Ala  
                                  485                      490                      495  
 Asp Leu Ala Gln Leu Phe Glu Val His Pro Asn Val Ala Ala Thr Ala  
 35                                   500                      505                      510  
 Val Arg Leu Ala Ala Arg Asp Ala Ala Arg Glu Val Ala Ala Cys Ser  
                                  515                      520                      525  
 Gln Leu Thr Ile Asn Ala Leu Arg Ser Pro Tyr Pro Ala His Pro Gly  
                                  530                      535                      540  
 40 Leu Leu Glu Leu Cys Val Ile Phe Phe Phe Glu Arg Val Leu Ala Phe  
 545                                   550                      555                      560  
 Leu Ile Glu Asn Gly Ala Arg Thr His Thr Gln Ala Gly Val Ala Gly  
                                  565                      570                      575

Pro Ala Ala Ala Leu Leu Asp Phe Thr Leu Arg Met Leu Pro Arg Lys  
 580 585 590  
 Thr Ala Val Gly Asp Phe Leu Ala Ser Thr Arg Met Ser Leu Ala Asp  
 595 600 605  
 5 Val Ala Ala His Arg Pro Leu Ile Gln His Val Leu Asp Glu Asn Ser  
 610 615 620  
 Gln Ile Gly Arg Leu Ala Lys Leu Val Leu Val Ala Arg Asp Val Ile  
 625 630 635 640  
 Arg Glu Thr Asp Ala Phe Tyr Gly Asp Leu Ala Asp Leu Asp Leu Gln  
 10 645 650 655  
 Leu Arg Ala Ala Pro Pro Ala Asn Leu Tyr Ala Arg Leu Gly Glu Trp  
 660 665 670  
 Leu Leu Glu Arg Ser Arg Ala His Pro Asn Thr Leu Phe Ala Pro Ala  
 675 680 685  
 15 Thr Pro Thr His Pro Glu Pro Leu Leu His Arg Ile Gln Ala Gln Phe  
 690 695 700  
 Arg Glu Glu Met Arg Val Glu Ala Glu Ala Arg Glu Met Arg Glu Ala  
 705 710 715 720  
 Leu Asp Arg Val Asp Ser Val Ser Gln Arg Ala Gly Pro Leu Thr Val  
 20 725 730 735  
 Met Pro Val Pro Ala Ala Pro Gly Ala Gly Gly Arg Ala Pro Cys Pro  
 740 745 750  
 Pro Ala Leu Gly Pro Glu Ala Ile Gln Ala Arg Leu Glu Asp Val Arg  
 755 760 765  
 25 Ile Gln Ala Arg Arg Ala Ile Glu Ser Ala Ile Lys Glu Tyr Phe His  
 770 775 780  
 Arg Gly Ala Val Tyr Ser Ala Lys Ala Leu Gln Ala Ser Asp Ser His  
 785 790 795 800  
 Asp Cys Arg Phe His Val Ala Ser Ala Ala Val Val Pro Met Val Gln  
 30 805 810 815  
 Leu Leu Glu Ser Leu Pro Ala Phe Asp Gln His Thr Arg Asp Val Ala  
 820 825 830  
 Gln Arg Ala Ala Leu Pro Pro Pro Pro Pro Leu Ala Thr Ser Pro Gln  
 835 840 845  
 35 Ala Ile Leu Leu Arg Asp Leu Leu Gln Arg Gly Gln Thr Leu Asp Ala  
 850 855 860  
 Pro Glu Asp Leu Ala Ala Trp Leu Ser Val Leu Thr Asp Ala Ala Thr  
 865 870 875 880  
 Gln Gly Leu Ile Glu Arg Lys Pro Leu Glu Glu Leu Ala Arg Ser Ile  
 40 885 890 895  
 His Gly Ile Asn Asp Gln Gln Ala Arg Arg Ser Ser Gly Leu Ala Glu  
 900 905 910  
 Leu Gln Arg Phe Asp Ala Leu Asp Ala Ala Gln Gln Leu Asp Ser Asp



	915	920	925
	Ala Ala Phe Val Pro Ala Thr Gly Pro Ala Pro Tyr Val Asp Gly Gly		
	930	935	940
	Gly Leu Ser Pro Glu Ala Thr Arg Met Ala Glu Asp Ala Leu Arg Gln		
5	945	950	955
	Ala Arg Ala Met Glu Ala Ala Lys Met Thr Ala Glu Leu Ala Pro Glu		
	965	970	975
	Ala Arg Ser Arg Leu Arg Glu Arg Ala His Ala Leu Glu Ala Met Leu		
	980	985	990
10	Asn Asp Ala Arg Glu Arg Ala Lys Val Ala His Asp Ala Arg Glu Lys		
	995	1000	1005
	Phe Leu His Lys Leu Gln Gly Val Leu Arg Pro Leu Pro Asp Phe Val		
	1010	1015	1020
	Gly Leu Lys Ala Cys Pro Ala Val Leu Ala Thr Leu Arg Ala Ser Leu		
15	1025	1030	1035
	Pro Ala Gly Trp Thr Asp Leu Ala Asp Ala Val Arg Gly Pro Pro Pro		
	1045	1050	1055
	Glu Val Thr Ala Ala Leu Arg Ala Asp Leu Trp Gly Leu Leu Gly Gln		
	1060	1065	1070
20	Tyr Arg Glu Ala Leu Glu His Pro Thr Pro Asp Thr Ala Thr Ala Gly		
	1075	1080	1085
	Leu His Pro Ala Phe Val Val Val Leu Lys Thr Leu Phe Ala Asp Ala		
	1090	1095	1100
	Pro Glu Thr Pro Val Leu Val Gln Phe Phe Ser Asp His Ala Pro Thr		
25	1105	1110	1115
	Ile Ala Lys Ala Val Ser Asn Ala Ile Asn Ala Gly Ser Ala Ala Val		
	1125	1130	1135
	Ala Thr Asp Ala Ala Thr Val Asp Ala Ala Val Arg Ala His Gly Ala		
	1140	1145	1150
30	Asp Ala Val Ser Ala Leu Gly Ala Ala Ala Arg Asp Pro Asp Leu Ser		
	1155	1160	1165
	Phe Leu Ala Ala Asp Ser Ala Ala Gly Tyr Val Lys Ala Thr Arg Leu		
	1170	1175	1180
	Ala Leu Glu Arg Ala Ile Asp Glu Leu Thr Thr Leu Gly Ser Ala Ala		
35	1185	1190	1195
	Ala Asp Leu Val Val Gln Ala Arg Arg Ala Cys Ala Gln Pro Glu Gly		
	1205	1210	1215
	Asp His Ala Ala Leu Ile Asp Ala Ala Ala Arg Ala Thr Thr Ala Ala		
	1220	1225	1230
40	Arg Glu Ser Leu Ala Gly His Glu Ala Gly Phe Gly Gly Leu Leu His		
	1235	1240	1245
	Ala Glu Gly Thr Ala Gly Asp His Ser Pro Ser Gly Arg Ala Leu Gln		
	1250	1255	1260

Glu Leu Gly Lys Val Ile Gly Ala Thr Arg Arg Arg Ala Asp Glu Leu  
 1265 1270 1275 128  
 Glu Ala Ala Val Ala Asp Leu Thr Ala Lys Met Ala Ala Gln Arg Arg  
 1285 1290 1295  
 5 Ser Ser Trp Ala Ala Gly Val Glu Ala Ala Leu Asp Arg Val Glu Asn  
 1300 1305 1310  
 Arg Ala Glu Phe Asp Val Val Glu Leu Arg Arg Leu Gln Ala Gly Thr  
 1315 1320 1325  
 His Gly Tyr Asn Pro Arg Asp Phe Arg Lys Arg Ala Glu Gln Ala Ala  
 10 1330 1335 1340  
 Asn Ala Glu Ala Val Thr Leu Ala Leu Asp Thr Ala Phe Ala Phe Asn  
 1345 1350 1355 136  
 Pro Tyr Thr Pro Glu Asn Gln Arg His Pro Met Leu Pro Pro Leu Ala  
 1365 1370 1375  
 15 Ala Ile His Arg Leu Gly Trp Ser Ala Ala Phe His Ala Ala Ala Glu  
 1380 1385 1390  
 Thr Tyr Ala Asp Met Phe Arg Val Asp Ala Glu Pro Leu Ala Arg Leu  
 1395 1400 1405  
 Leu Arg Ile Ala Glu Gly Leu Leu Glu Met Ala Gln Ala Gly Asp Gly  
 20 1410 1415 1420  
 Phe Ile Asp Tyr His Glu Ala Val Gly Arg Leu Ala Asp Asp Met Thr  
 1425 1430 1435 144  
 Ser Val Pro Gly Leu Arg Arg Tyr Val Pro Phe Phe Gln His Gly Tyr  
 1445 1450 1455  
 25 Ala Asp Tyr Val Glu Leu Arg Asp Arg Leu Asp Ala Ile Arg Ala Asp  
 1460 1465 1470  
 Val His Arg Ala Leu Gly Gly Val Pro Leu Asp Leu Ala Ala Ala Ala  
 1475 1480 1485  
 Glu Gln Ile Ser Ala Ala Arg Asn Asp Pro Glu Ala Thr Ala Glu Leu  
 30 1490 1495 1500  
 Val Arg Thr Gly Val Thr Leu Pro Cys Pro Ser Glu Asp Ala Leu Val  
 1505 1510 1515 152  
 Ala Cys Ala Ala Ala Leu Glu Arg Val Asp Gln Ser Pro Val Lys Asn  
 1525 1530 1535  
 35 Thr Ala Tyr Ala Glu Tyr Val Ala Phe Val Thr Arg Gln Asp Thr Ala  
 1540 1545 1550  
 Glu Thr Lys Asp Ala Val Val Arg Ala Lys Gln Gln Arg Ala Glu Ala  
 1555 1560 1565  
 Thr Glu Arg Val Met Ala Gly Leu Arg Glu Ala Ala Arg Glu Arg Arg  
 40 1570 1575 1580  
 Ala Gln Ile Glu Ala Glu Gly Leu Ala Asn Leu Lys Thr Met Leu Lys  
 1585 1590 1595 160  
 Val Val Ala Val Pro Ala Thr Val Ala Lys Thr Leu Asp Gln Ala Arg

	1605	1610	1615
	Ser Val Ala Glu Ile Ala Asp Gln Val Glu Val Leu Leu Asp Gln Thr		
	1620	1625	1630
	Glu Lys Thr Arg Glu Leu Asp Val Pro Ala Val Ile Trp Leu Glu His		
5	1635	1640	1645
	Ala Gln Arg Thr Phe Glu Thr His Pro Leu Ser Ala Arg Asp Gly Pro		
	1650	1655	1660
	Gly Pro Leu Ala Arg His Ala Gly Arg Leu Gly Ala Leu Phe Asp Thr		
	1665	1670	1675
10	Arg Arg Arg Val Asp Ala Leu Arg Arg Ser Leu Glu Glu Ala Glu Ala		
	1685	1690	1695
	Glu Trp Asp Glu Val Trp Gly Arg Phe Gly Arg Val Arg Gly Gly Ala		
	1700	1705	1710
	Trp Lys Ser Pro Glu Gly Phe Arg Ala Met His Glu Gln Leu Arg Ala		
15	1715	1720	1725
	Leu Gln Asp Thr Thr Asn Thr Val Ser Gly Leu Arg Ala Gln Pro Ala		
	1730	1735	1740
	Tyr Glu Arg Leu Ser Ala Arg Tyr Gln Gly Val Leu Gly Ala Lys Gly		
	1745	1750	1755
20	Ala Glu Arg Ala Glu Ala Val Glu Glu Leu Gly Ala Arg Val Thr Lys		
	1765	1770	1775
	His Thr Ala Leu Cys Ala Arg Leu Arg Asp Glu Val Val Arg Arg Val		
	1780	1785	1790
	Pro Trp Glu Met Asn Phe Asp Ala Leu Gly Arg Leu Leu Ala Glu Phe		
25	1795	1800	1805
	Asp Ala Ala Ala Ala Asp Leu Ala Pro Trp Ala Val Glu Glu Phe Arg		
	1810	1815	1820
	Gly Ala Arg Glu Leu Ile Gln Tyr Arg Met Gly Ser Ala Tyr Ala Arg		
	1825	1830	1835
30	Ala Gly Gly Gln Thr Xaa Xaa Xaa Xaa		
	1845	1850	

## (2) INFORMATION FOR SEQ ID NO:121:

35

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1100 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

40

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

	Met	Ser	Asp	Ser	Ala	Leu	Gln	Val	Pro	Ala	Pro	Ala	Gly	Met	Thr	Pro
	1				5					10					15	
	Pro	Ser	Ala	Pro	Pro	Pro	Asn	Gly	Pro	Leu	Gln	Val	Leu	Leu	Gly	Ser
5				20					25					30		
	Leu	Thr	Asn	Leu	Arg	Arg	Pro	Pro	Ser	Pro	Ser	Ser	Glu	Pro	Ala	Gly
			35					40					45			
	Ser	Ala	Asp	Glu	Pro	Ala	Phe	Leu	Ser	Ala	Ala	Lys	Leu	Arg	Ala	Ala
			50				55					60				
10	Thr	Ala	Ala	Phe	Leu	Leu	Ser	Gly	Ala	Ala	Val	Gly	Pro	Ala	Glu	Ala
	65					70					75				80	
	Arg	Ala	Cys	Trp	His	Pro	Leu	Leu	Glu	Gln	Leu	Cys	Ala	Leu	His	Arg
					85					90					95	
	Ala	His	Gly	Leu	Pro	Glu	Thr	Ala	Leu	Leu	Ala	Glu	Asn	Leu	Pro	Gly
15				100					105					110		
	Leu	Leu	Val	His	Arg	Met	Ala	Val	Pro	Glu	Thr	Pro	Glu	Ala	Ala	Phe
			115					120					125			
	Arg	Glu	Met	Asp	Val	Ile	Lys	Asp	Thr	Val	Leu	Ala	Ile	Thr	Gly	Ser
		130					135				140					
20	Asp	Thr	Thr	His	Ala	Leu	Glu	Ala	Ala	Gly	Leu	Arg	Thr	Thr	Ala	Ala
	145					150					155				160	
	Leu	Gly	Pro	Val	Arg	Val	Arg	Gln	Cys	Ala	Val	Glu	Trp	Ile	Asp	Arg
					165					170					175	
	Trp	Arg	Thr	Val	Thr	Gln	Ser	Cys	Leu	Ala	Met	Asn	Pro	Arg	Thr	Ser
25				180					185					190		
	Leu	Glu	Ala	Leu	Gly	Glu	Met	Ser	Leu	Lys	Met	Ser	Pro	Val	Pro	Leu
			195					200				205				
	Gly	Gln	Pro	Gly	Ala	Asn	Leu	Thr	Thr	Pro	Ala	Tyr	Ser	Leu	Leu	Phe
		210					215					220				
30	Pro	Ser	Pro	Ile	Val	Gln	Glu	Gly	Leu	Arg	Phe	Leu	Ala	Leu	Val	Ser
	225					230					235				240	
	Asn	Trp	Val	Thr	Leu	Phe	Ser	Ala	His	Leu	Gln	Arg	Ile	Asp	Asp	Ala
					245					250				255		
	Ala	Leu	Thr	Pro	Leu	Thr	Arg	Ala	Leu	Phe	Thr	Leu	Ala	Leu	Val	Asp
35				260					265					270		
	Glu	Tyr	Leu	Thr	Thr	Pro	Asp	Arg	Gly	Ala	Val	Val	Pro	Pro	Pro	Leu
			275					280					285			
	Leu	Ala	Gln	Phe	Gln	His	Thr	Val	Arg	Glu	Ile	Asp	Pro	Ala	Ile	Met
		290					295					300				
40	Ile	Pro	Pro	Leu	Glu	Ala	Thr	Lys	Met	Val	Arg	Ser	Arg	Glu	Glu	Val
	305					310					315				320	
	Arg	Val	Ser	Thr	Ala	Leu	Ser	Arg	Val	Ser	Pro	Arg	Ser	Ala	Cys	Ala
				325						330				335		

Pro Pro Gly Thr Leu Met Ala Arg Val Arg Thr Asp Ala Ala Val Phe  
 340 345 350  
 Asp Pro Asp Val Pro Phe Leu Ser Ala Ser Ala Ile Phe Arg Pro Ala  
 355 360 365  
 5 Val Thr Gly Leu Leu Gln Leu Gly Glu Pro Pro Ser Ala Gly Ala Gln  
 370 375 380  
 Gln Arg Leu Leu Ala Leu Leu Gln Gln Thr Trp Ala Leu Val Gln Asn  
 385 390 395 400  
 Ser Asn Ser Pro Ser Val Val Ile Asn Thr Leu Thr Asp Ala Gly Phe  
 10 405 410 415  
 Thr Pro Ala His Cys Thr Gln Tyr Ile Ser Ala Leu Glu Gly Phe Leu  
 420 425 430  
 Val Ala Gly Val Pro Ala Arg Thr Pro Pro Gly His Gly Leu Ser Glu  
 435 440 445  
 15 Ile Gln Gln Leu Phe Gly Cys Ile Ala Gly Ala Asn Val Phe Gly Leu  
 450 455 460  
 Ala Arg Glu Tyr Gly His Tyr Ala Gly Tyr Val Lys Thr Phe Arg Arg  
 465 470 475 480  
 Ile Gln Gly Ala Ser Glu His Thr His Gly Arg Leu Cys Glu Ala Val  
 20 485 490 495  
 Gly Leu Ser Gly Gly Val Leu Ser Gln Thr Leu Ala Arg Ile Met Gly  
 500 505 510  
 Pro Ala Val Pro Thr Glu His Leu Ala Ser Leu Arg Arg Thr Leu Val  
 515 520 525  
 25 Gly Glu Phe Glu Thr Ala Glu Arg Arg Phe Ser Ala Gly Gln Pro Ser  
 530 535 540  
 Leu Leu Arg Glu Thr Ala Leu Ile Trp Leu Asp Val Tyr Gly Gln Thr  
 545 550 555 560  
 His Trp Asp Leu Thr Pro Thr Thr Pro Ala Thr Pro Leu Ser Ala Leu  
 30 565 570 575  
 Leu Pro Val Gly Pro Pro Ser His Ala Pro Ser Val His Leu Ala Ala  
 580 585 590  
 Ala Thr Lys Ile Arg Phe Pro Ala Leu Glu Gly Ile His Pro Asn Val  
 595 600 605  
 35 Leu Ala Asp Pro Gly Phe Val Pro Tyr Val Leu Ala Leu Val Val Gly  
 610 615 620  
 Asp Ala Leu Arg Ala Thr Cys Asn Ala Ala Tyr Leu Pro Arg Pro Ile  
 625 630 635 640  
 Glu Phe Ala Leu Arg Val Leu Ala Trp Ala Arg Asp Phe Gly Leu Gly  
 40 645 650 655  
 Tyr Leu Pro Thr Val Glu Gly His Arg Thr Lys Leu Gly Ala Leu Ile  
 660 665 670  
 Thr Leu Leu Glu Pro Ala Thr Arg Ala Gly Val Gly Pro Thr Met Gln  
 366

	675		680		685	
	Met Ala Asp Asn Ile Glu Gln Leu Leu Arg Glu Leu Tyr Val Ile Arg					
	690		695		700	
5	Ala Val Glu Gln Leu Arg Pro Ala Val Gln Leu Pro Pro Pro Gln Pro					
	705		710		715	720
	Pro Glu Val Gly Ser Ser Leu Leu Leu Ile Ser Met Tyr Ala Arg Val					
		725		730		735
	Leu Gln Glu Leu Ala Glu Arg Ala Asp Pro Leu Val Arg Gln Leu Glu					
		740		745		750
10	Asp Ala Ile Val Leu Leu Arg Leu His Met Arg Thr Leu Ala Ala Phe					
		755		760		765
	Phe Glu Cys Arg Phe Glu Ser Asp Gly His Arg Leu Tyr Ala Val Val					
		770		775		780
	Ala Asp Ala His Glu Arg Leu Gly Pro Trp Arg Pro Glu Ala Met Gly					
15		785		790		800
	Asp Ala Val Ser Gln Tyr Cys Gly Met Tyr His Asp Ala Lys Arg Ala					
		805		810		815
	Leu Val Ala Ser Leu Ala Gly Leu Arg Ser Val Val Thr Glu Thr Thr					
		820		825		830
20	Ala His Leu Gly Val Cys Asp Glu Leu Ala Ala Gln Val Ser His Glu					
		835		840		845
	Gly Asn Val Leu Ala Val Val Arg Arg Glu Ile His Gly Phe Leu Ala					
		850		855		860
	Ile Val Ser Gly Ile His Ala Arg Ala Ser Lys Leu Met Ser Gly Asp					
25		865		870		880
	Gln Val Pro Gly Phe Cys Tyr Met Ser Gln Phe Leu Ala Arg Trp Arg					
		885		890		895
	Arg Leu Ser Ala Gly Tyr Gln Ala Ala Arg Ala Ala Thr Gly Pro Glu					
		900		905		910
30	Arg Val Ala Glu Phe Val Gln Glu Leu His Asp Thr Trp Lys Gly Leu					
		915		920		925
	Gln Thr Glu Arg Ala Leu Val Val Ala Pro Phe Ala Ser Ser Ala Asp					
		930		935		940
	Gln Arg Thr Ala Ala Ile Gln Glu Val Met Ala His Ala Thr Glu Asp					
35		945		950		960
	Ala Pro Pro Ser Pro Ala Ala Asp Leu Val Val Leu Thr Asn Arg His					
		965		970		975
	Asp Leu Gly Ala Trp Gly Asp Tyr Ser Leu Gly Pro Leu Gly Gln Pro					
		980		985		990
40	Thr Val Val Pro Asp Ser Val Asp Leu Ser Pro Gln Gly Leu Ala Ala					
		995		1000		1005
	Thr Leu Ser Met Asp Trp Leu Leu Ile Asn Glu Leu Leu Gln Val Thr					
		1010		1015		1020

```

      Asp Gly Val Phe Arg Ala Ser Ala Phe Arg Pro Ser Ala Gly Pro Glu
1025                      1030                      1035                      104
      Ala Pro Gly Asp Leu Glu Ala Gln Asp Ala Gly Gly Ser Thr Pro Glu
                      1045                      1050                      1055
5    Pro Thr Thr Pro Gly Pro Gln Asp Thr Gln Ala Arg Ala Pro Ser Trp
                      1060                      1065                      1070
      Ala Gly Arg Glu Thr Val Pro Trp Pro Asn Thr Pro Val Glu Asp Asp
                      1075                      1080                      1085
      Glu Met Thr Pro Gln Glu Thr Pro Pro Val His Pro
10    1090                      1095                      1100

```

## (2) INFORMATION FOR SEQ ID NO:122:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 641 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

```

      Val Glu Arg Thr Gly Gly Ser Cys Arg Arg Ala Pro Gly Pro Gly Ala
25    1          5          10          15
      Arg Cys Pro Thr Trp Arg Pro Ala Cys Ala Leu Gly Asp Ala Ala Arg
                      20          25          30
      Arg Pro Arg Ala Gln Thr Gly Met Thr Ala Ala Ala Leu Tyr Gly Gly
                      35          40          45
30    Ala Lys Tyr Arg Pro Gly Thr Leu Arg Asn Pro Gly Arg Val Ala Ser
                      50          55          60
      Thr Pro Arg Arg Arg Gly Val Leu Tyr Gly Ala Leu Cys Pro Gly Ile
65          70          75          80
      Pro Phe Val Gly Ser Gly Pro Gly Ala Val Gly Trp Glu Cys Val Cys
35          85          90          95
      Val Gly Gly Gly Arg Arg Asp Gly Gly Pro Asp Gln Val Tyr Arg Gly
                      100          105          110
      Arg Ser Val Gly Arg Pro Asn Arg Pro Phe Lys His Leu Arg Met His
                      115          120          125
40    Arg Pro Ser Gln Ser Asp Thr Gly Thr His Gln Arg Arg Lys Pro Pro
                      130          135          140
      Ser Pro Val Arg Val Arg Val Phe Ser Gly Gly Val Phe Phe Leu Ser
145          150          155          160

```

Ala Leu Leu Pro Pro His Leu His His Pro Pro Pro Thr Trp Leu Ala  
 165 170 175  
 Ile Gly Gly Lys Thr Met Lys Thr Lys Pro Leu Pro Thr Ala Pro Met  
 180 185 190  
 5 Ala Trp Ala Glu Ser Ala Val Glu Thr Thr Thr Ser Pro Arg Glu Leu  
 195 200 205  
 Ala Gly His Ala Pro Leu Arg Arg Val Leu Arg Pro Pro Ile Ala Arg  
 210 215 220  
 Arg Asp Gly Pro Val Leu Leu Gly Asp Arg Ala Pro Arg Arg Thr Ala  
 10 225 230 235 240  
 Ser Thr Met Trp Leu Leu Gly Ile Asp Pro Ala Glu Ser Ser Pro Gly  
 245 250 255  
 Thr Arg Ala Thr Arg Asp Asp Thr Glu Gln Ala Val Asp Lys Ile Leu  
 260 265 270  
 15 Arg Gly Ala Arg Arg Ala Gly Gly Leu Thr Val Pro Gly Ala Pro Arg  
 275 280 285  
 Tyr His Leu Thr Arg Gln Val Thr Leu Thr Asp Leu Cys Gln Pro Asn  
 290 295 300  
 Ala Glu Arg Ala Gly Ala Leu Leu Leu Ala Leu Arg His Pro Thr Asp  
 20 305 310 315 320  
 Leu Pro His Leu Ala Arg His Arg Ala Pro Pro Gly Arg Gln Thr Glu  
 325 330 335  
 Arg Leu Ala Glu Ala Trp Gly Gln Leu Leu Glu Ala Ser Ala Leu Gly  
 340 345 350  
 25 Ser Gly Arg Ala Glu Ser Gly Cys Ala Arg Ala Gly Leu Val Ser Phe  
 355 360 365  
 Asn Phe Leu Val Ala Ala Cys Ala Ala Ala Tyr Asp Ala Arg Asp Ala  
 370 375 380  
 Ala Glu Ala Val Arg Ala His Ile Thr Thr Asn Tyr Gly Gly Thr Arg  
 30 385 390 395 400  
 Ala Gly Ala Arg Leu Asp Arg Phe Ser Glu Cys Leu Arg Ala Met Val  
 405 410 415  
 His Thr His Val Phe Phe Val Met Arg Phe Phe Gly Gly Leu Val Ser  
 420 425 430  
 35 Trp Val Thr Gln Asp Glu Leu Ala Ser Val Thr Ala Val Cys Ser Gly  
 435 440 445  
 Pro Gln Glu Ala Thr His Thr Gly His Pro Gly Arg Pro Cys Ser Ala  
 450 455 460  
 Val Thr Ile Pro Ala Cys Ala Phe Val Asp Leu Asp Ala Glu Leu Cys  
 40 465 470 475 480  
 Leu Gly Gly Pro Gly Ala Ala Phe Leu Tyr Leu Val Phe Tyr Gln Cys  
 485 490 495  
 Arg Asp Gln Glu Leu Cys Cys Val Tyr Val Val Lys Ser Gln Leu Pro



500 505 510  
 Pro Arg Gly Leu Glu Ala Ala Leu Glu Arg Leu Phe Gly Arg Leu Arg  
 515 520 525  
 Ile Thr Asn Thr Ile His Gly Ala Glu Asp Met Thr Pro Pro Pro Pro  
 5 530 535 540  
 Asn Arg Asn Val Asp Phe Pro Leu Ala Val Leu Ala Ala Ser Ser Gln  
 545 550 555 560  
 Ser Pro Arg Cys Ser Ala Ser Gln Val Thr Asn Pro Gln Phe Val Asp  
 565 570 575  
 10 Arg Leu Tyr Arg Trp Gln Pro Asp Leu Arg Gly Arg Pro Thr Ala Arg  
 580 585 590  
 Thr Cys Thr Tyr Ala Ala Phe Ala Glu Leu Gly Val Met Pro Asp Asn  
 595 600 605  
 Ser Pro Arg Cys Leu His Arg Thr Glu Arg Phe Gly Ala Val Gly Val  
 15 610 615 620  
 Pro Val Val Ile Gly Val Val Trp Arg Pro Gly Gly Trp Arg Ala Cys  
 625 630 635 640  
 Ala

20

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:

25

- (A) LENGTH: 1160 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Val Ile Arg Arg Pro Val Arg Pro Phe Gly Arg Thr Ala His Pro Ala  
 1 5 10 15  
 35 Ser His Gly Pro Ala Ala Val Ser Val His Arg Val Arg Ala Thr Val  
 20 25 30  
 Thr Leu Val Pro Met Ala Asn Arg Pro Ala Ala Ser Ala Gly Ala Arg  
 35 40 45  
 Ser Pro Ser Gln Glu Pro Arg Glu Pro Glu Val Ala Pro Pro Gly Gly  
 40 50 55 60  
 Asp His Val Phe Cys Arg Lys Val Ser Gly Val Met Val Leu Ser Ser  
 65 70 75 80  
 Asp Pro Pro Gly Pro Ala Ala Tyr Arg Ile Ser Asp Ser Ser Phe Val

370

				85					90					95					
				Gln	Cys	Gly	Ser	Asn	Cys	Ser	Met	Ile	Ile	Asp	Gly	Asp	Val	Arg	His
				100								105					110		
				Leu	Arg	Asp	Leu	Glu	Gly	Ala	Thr	Ser	Thr	Gly	Ala	Phe	Val	Ala	Ile
5				115							120				125				
				Ser	Asn	Val	Ala	Ala	Gly	Gly	Asp	Gly	Arg	Thr	Ala	Val	Val	Gly	Gly
				130							135				140				
				Thr	Ser	Gly	Pro	Ser	Ala	Thr	Thr	Ser	Val	Gly	Thr	Gln	Thr	Ser	Gly
				145					150					155				160	
10				Glu	Phe	Leu	His	Gly	Asn	Pro	Arg	Thr	Pro	Glu	Pro	Gln	Gly	Pro	Gln
								165					170				175		
				Ala	Val	Pro	Pro	Pro	Pro	Pro	Pro	Phe	Pro	Trp	Gly	His	Glu	Cys	
								180				185				190			
				Cys	Ala	Arg	Arg	Asp	Arg	Gly	Ala	Glu	Lys	Asp	Val	Gly	Ala	Ala	Glu
15				195							200				205				
				Ser	Trp	Ser	Asp	Gly	Pro	Ser	Ser	Asp	Ser	Glu	Thr	Glu	Asp	Ser	Asp
				210							215				220				
				Ser	Ser	Asp	Glu	Asp	Thr	Gly	Ser	Gly	Ser	Glu	Thr	Leu	Ser	Arg	Ser
				225					230					235				240	
20				Ser	Ser	Ile	Trp	Ala	Ala	Gly	Ala	Thr	Asp	Asp	Asp	Asp	Ser	Asp	Ser
								245					250				255		
				Asp	Ser	Arg	Ser	Asp	Asp	Ser	Val	Gln	Pro	Asp	Val	Val	Val	Arg	Arg
								260				265				270			
				Arg	Trp	Ser	Asp	Gly	Pro	Ala	Pro	Val	Ala	Phe	Pro	Lys	Pro	Arg	Arg
25				275							280				285				
				Pro	Gly	Asp	Ser	Pro	Gly	Asn	Pro	Gly	Leu	Gly	Ala	Gly	Thr	Gly	Pro
				290						295					300				
				Gly	Ser	Ala	Thr	Asp	Pro	Arg	Ala	Ser	Ala	Asp	Ser	Asp	Ser	Ala	Ala
				305					310					315				320	
30				His	Ala	Ala	Ala	Pro	Gln	Ala	Asp	Val	Ala	Pro	Val	Leu	Asp	Ser	Gln
								325					330				335		
				Pro	Thr	Val	Gly	Thr	Asp	Pro	Gly	Tyr	Pro	Val	Pro	Leu	Glu	Leu	Thr
								340				345				350			
				Pro	Glu	Asn	Ala	Glu	Ala	Val	Ala	Arg	Phe	Leu	Gly	Asp	Ala	Val	Asp
35				355							360				365				
				Arg	Glu	Pro	Ala	Leu	Met	Leu	Glu	Tyr	Phe	Cys	Arg	Cys	Ala	Arg	Glu
				370						375					380				
				Glu	Ser	Lys	Arg	Val	Pro	Pro	Arg	Thr	Phe	Gly	Ser	Ala	Pro	Arg	Leu
				385					390					395				400	
40				Thr	Glu	Asp	Asp	Phe	Gly	Leu	Leu	Asn	Tyr	Ala	Glu	Met	Arg	Arg	Leu
								405					410				415		
				Cys	Leu	Asp	Leu	Pro	Pro	Val	Pro	Pro	Asn	Ala	Tyr	Thr	Pro	Tyr	His
								420				425				430			

Leu Arg Glu Tyr Ala Thr Arg Leu Val Asn Gly Phe Lys Pro Leu Val  
 435 440 445  
 Arg Arg Ser Ala Arg Leu Tyr Arg Ile Leu Gly Ile Leu Val His Leu  
 450 455 460  
 5 Arg Ile Arg Thr Arg Glu Ala Ser Phe Glu Glu Trp Met Arg Ser Lys  
 465 470 475 480  
 Glu Val Asp Leu Asp Phe Gly Leu Thr Glu Arg Leu Arg Glu His Glu  
 485 490 495  
 Ala Gln Leu Met Ile Leu Ala Gln Ala Leu Asn Pro Tyr Asp Cys Leu  
 10 500 505 510  
 Ile His Ser Thr Pro Asn Thr Leu Val Glu Arg Gly Leu Gln Ser Ala  
 515 520 525  
 Leu Lys Tyr Glu Glu Phe Tyr Leu Lys Arg Phe Gly Gly His Tyr Met  
 530 535 540  
 15 Glu Ser Val Phe Gln Met Tyr Thr Arg Ile Ala Gly Phe Leu Ala Cys  
 545 550 555 560  
 Arg Ala Thr Arg Gly Met Arg His Ile Ala Leu Gly Arg Gln Gly Ser  
 565 570 575  
 Trp Trp Glu Met Phe Lys Phe Phe Phe His Arg Leu Tyr Asp His Gln  
 20 580 585 590  
 Ile Val Pro Ser Thr Pro Ala Met Leu Asn Leu Gly Thr Arg Asn Tyr  
 595 600 605  
 Tyr Thr Ser Ser Cys Tyr Leu Val Asn Pro Gln Ala Thr Thr Asn Gln  
 610 615 620  
 25 Ala Thr Leu Arg Ala Ile Thr Gly Asn Val Ser Ala Ile Leu Ala Arg  
 625 630 635 640  
 Asn Gly Gly Ile Gly Leu Cys Met Gln Ala Phe Asn Asp Asp Gly Thr  
 645 650 655  
 Ala Ser Ile Met Pro Ala Leu Lys Val Leu Asp Ser Leu Val Ala Ala  
 30 660 665 670  
 His Asn Lys Gln Ser Trp Thr Gly Ala Cys Val Tyr Leu Glu Pro Trp  
 675 680 685  
 His Ser Asp Val Arg Ala Val Leu Arg Met Lys Gly Val Leu Ala Gly  
 690 695 700  
 35 Glu Glu Ala Gln Arg Cys Asp Asn Ile Phe Ser Ala Leu Trp Met Pro  
 705 710 715 720  
 Asp Leu Phe Phe Lys Arg Leu Ile Arg His Leu Asp Gly Glu Lys Asn  
 725 730 735  
 Val Thr Trp Ser Leu Phe Asp Arg Asp Thr Ser Met Ser Leu Ala Asp  
 40 740 745 750  
 Phe His Gly Glu Glu Phe Glu Lys Leu Tyr Glu His Leu Glu Ala Met  
 755 760 765  
 Gly Phe Gly Glu Thr Ile Pro Ile Gln Asp Leu Ala Tyr Ala Ile Val  
 372

	770		775		780
	Arg Ser Ala Ala Thr Thr Gly Ser Pro Phe Ile Met Phe Lys Asp Ala				
	785		790		795 800
5	Val Asn Arg His Tyr Ile Tyr Asp Thr Gln Gly Ala Ala Ile Ala Gly				
		805		810	815
	Ser Asn Leu Cys Thr Glu Ile Val His Pro Ser Ser Lys Arg Ser Ser				
		820		825	830
	Gly Val Cys Asn Leu Gly Ser Val Asn Leu Ala Arg Cys Val Ser Arg				
		835		840	845
10	Arg Thr Phe Asp Phe Gly Met Leu Arg Asp Ala Val Gln Ala Cys Val				
		850		855	860
	Leu Met Val Asn Ile Met Ile Asp Ser Thr Leu Gln Pro Thr Pro Gln				
		865		870	875 880
	Cys Arg His Asp Asn Leu Arg Ser Met Gly Ile Gly Met Gln Gly Leu				
15		885		890	895
	His Thr Ala Cys Leu Lys Met Gly Leu Asp Leu Glu Ser Ala Glu Phe				
		900		905	910
	Arg Asp Leu Asn Thr His Ile Ala Glu Val Met Leu Leu Ala Ala Met				
		915		920	925
20	Lys Thr Ser Asn Ala Leu Cys Val Arg Gly Ala Arg Pro Phe Ser His				
		930		935	940
	Phe Lys Arg Ser Met Tyr Arg Ala Gly Arg Phe His Trp Glu Arg Phe				
		945		950	955 960
	Ser Asn Asp Arg Tyr Glu Gly Glu Trp Glu Met Leu Arg Gln Ser Met				
25		965		970	975
	Met Lys His Gly Leu Arg Asn Ser Gln Phe Ile Ala Leu Met Pro Thr				
		980		985	990
	Ala Ala Ser Ala Gln Ile Ser Asp Val Ser Glu Gly Phe Ala Pro Leu				
		995		1000	1005
30	Phe Thr Asn Leu Phe Ser Lys Val Thr Arg Asp Gly Glu Thr Leu Arg				
		1010		1015	1020
	Pro Asn Thr Leu Leu Leu Lys Glu Leu Glu Arg Thr Phe Gly Gly Lys				
		1025		1030	1035 104
	Arg Leu Leu Asp Ala Met Asp Gly Leu Glu Ala Lys Gln Trp Ser Val				
35		1045		1050	1055
	Ala Gln Ala Leu Pro Cys Leu Asp Pro Ala His Pro Leu Arg Arg Phe				
		1060		1065	1070
	Lys Thr Ala Phe Asp Tyr Asp Gln Glu Leu Leu Ile Asp Leu Cys Ala				
		1075		1080	1085
40	Asp Arg Ala Pro Tyr Val Asp His Ser Gln Ser Met Thr Leu Tyr Val				
		1090		1095	1100
	Thr Glu Lys Ala Asp Gly Thr Leu Pro Ala Ser Thr Leu Val Arg Leu				
		1105		1110	1115 112

Leu Val His Ala Tyr Lys Arg Gly Leu Lys Thr Gly Met Tyr Tyr Cys  
 1125 1130 1135  
 Lys Val Arg Lys Ala Thr Asn Ser Gly Val Phe Ala Gly Asp Asp Asn  
 1140 1145 1150  
 5 Ile Val Cys Thr Ser Cys Ala Leu  
 1155 1160

## (2) INFORMATION FOR SEQ ID NO:124:

## 10 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 333 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

20 Met Asp Pro Ala Val Ser Pro Ala Ser Thr Asp Pro Leu Asp Thr His  
 1 5 10 15  
 Ala Ser Gly Ala Gly Ala Ala Pro Ile Pro Val Cys Pro Thr Pro Glu  
 20 25 30  
 Arg Tyr Phe Tyr Thr Ser Gln Cys Pro Asp Ile Asn His Leu Arg Ser  
 25 35 40 45  
 Leu Ser Ile Leu Asn Arg Trp Leu Glu Thr Glu Leu Val Phe Val Gly  
 50 55 60  
 Asp Glu Glu Asp Val Ser Lys Leu Ser Glu Gly Glu Leu Gly Phe Tyr  
 65 70 75 80  
 30 Arg Phe Leu Phe Ala Phe Leu Ser Ala Ala Asp Asp Leu Val Thr Glu  
 85 90 95  
 Asn Leu Gly Gly Leu Ser Gly Leu Phe Glu Gln Lys Asp Ile Leu His  
 100 105 110  
 Tyr Tyr Val Glu Gln Glu Cys Ile Glu Val Val His Ser Arg Val Tyr  
 115 120 125  
 35 Asn Ile Ile Gln Leu Val Leu Phe His Asn Asn Asp Gln Ala Arg Arg  
 130 135 140  
 Ala Tyr Val Ala Arg Thr Ile Asn His Pro Ala Ile Arg Val Lys Val  
 145 150 155 160  
 40 Asp Trp Leu Glu Ala Arg Val Arg Glu Cys Asp Ser Ile Pro Glu Lys  
 165 170 175  
 Phe Ile Leu Met Ile Leu Ile Glu Gly Val Phe Phe Ala Ala Ser Phe  
 180 185 190

Ala Ala Ile Ala Tyr Leu Arg Thr Asn Asn Leu Leu Arg Val Thr Cys  
 195 200 205  
 Gln Ser Asn Asp Leu Ile Ser Arg Asp Glu Ala Val His Thr Thr Ala  
 210 215 220  
 5 Ser Cys Tyr Ile Tyr Asn Asn Tyr Leu Gly Gly His Ala Lys Pro Glu  
 225 230 235 240  
 Ala Ala Arg Val Tyr Arg Leu Phe Arg Glu Ala Val Asp Ile Glu Ile  
 245 250 255  
 Gly Phe Ile Arg Ser Gln Ala Pro Thr Asp Ser Ser Ile Leu Ser Pro  
 10 260 265 270  
 Gly Ala Ala Ile Glu Asn Tyr Val Arg Phe Ser Ala Asp Arg Leu Leu  
 275 280 285  
 Gly Leu Ile His Met Gln Pro Lys Ala Pro Ala Pro Asp Ala Ser Phe  
 290 295 300  
 15 Pro Leu Ser Leu Met Ser Thr Asp Lys His Thr Asn Phe Phe Glu Cys  
 305 310 315 320  
 Arg Ser Thr Ser Tyr Ala Gly Ala Val Val Asn Asp Leu  
 325 330

20 (2) INFORMATION FOR SEQ ID NO:125:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 357 amino acids  
 (B) TYPE: amino acid  
 25 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

Met Arg Arg Arg Gly His Ala Phe Ala Pro Gly Asp Arg Gly Thr Arg  
 1 5 10 15  
 Ala Ala Gly Pro Gly Pro Ala Ala Pro Trp Gly Ala Pro Ser Lys Pro  
 20 25 30  
 35 Ala Leu Arg Leu Ala His Leu Phe Cys Ile Arg Val Leu Arg Ala Leu  
 35 40 45  
 Gly Tyr Ala Tyr Ile Asn Ser Gly Gln Leu Glu Ala Asp Asp Ala Cys  
 50 55 60  
 40 Ala Asn Leu Tyr His Thr Asn Thr Val Ala Tyr Val His Thr Thr Asp  
 65 70 75 80  
 Thr Asp Leu Leu Leu Met Gly Cys Asp Ile Val Leu Asp Ile Ser Thr  
 85 90 95

Gly Tyr Ile Pro Thr Ile His Cys Arg Asp Leu Leu Gln Tyr Phe Lys  
                   100                  105                  110  
 Met Ser Tyr Pro Gln Phe Leu Ala Leu Phe Val Arg Cys His Thr Asp  
                   115                  120                  125  
 5 Leu His Pro Asn Asn Thr Tyr Ala Ser Val Glu Asp Val Leu Arg Glu  
                   130                  135                  140  
 Cys His Trp Thr Ala Pro Ser Arg Ser Gln Ala Arg Arg Ala Ala Arg  
 145                  150                  155                  160  
 Arg Glu Arg Ala Asn Ser Arg Ser Leu Glu Ser Met Pro Thr Leu Thr  
 10                  165                  170                  175  
 Ala Ala Pro Val Gly Leu Glu Thr Arg Ile Ser Trp Thr Glu Ile Leu  
                   180                  185                  190  
 Ala Gln Gln Ile Ala Gly Glu Asp Asp Tyr Glu Glu Asp Pro Pro Leu  
                   195                  200                  205  
 15 Gln Pro Pro Asp Val Ala Gly Gly Pro Arg Asp Gly Ala Arg Ser Ser  
                   210                  215                  220  
 Ser Ser Glu Ile Leu Thr Pro Pro Glu Leu Val Gln Val Pro Asn Ala  
 225                  230                  235                  240  
 Gln Arg Val Ala Glu His Arg Gly Tyr Val Ala Gly Arg Arg Arg His  
 20                  245                  250                  255  
 Val Ile His Asp Ala Pro Glu Ala Leu Asp Trp Leu Pro Asp Pro Met  
                   260                  265                  270  
 Thr Ile Ala Glu Leu Val Glu His Arg Tyr Val Lys Tyr Val Ile Ser  
                   275                  280                  285  
 25 Leu Ile Ser Pro Lys Glu Arg Gly Pro Trp Thr Leu Leu Lys Arg Leu  
                   290                  295                  300  
 Pro Ile Tyr Gln Asp Leu Arg Asp Glu Asp Leu Ala Arg Ser Ile Val  
 305                  310                  315                  320  
 Thr Arg His Ile Thr Ala Pro Asp Ile Ala Asp Arg Phe Leu Ala Gln  
 30                  325                  330                  335  
 Leu Trp Ala His Ala Pro Pro Pro Ala Phe Tyr Lys Asp Val Leu Ala  
                   340                  345                  350  
 Lys Phe Trp Asp Glu  
                   355

35

## (2) INFORMATION FOR SEQ ID NO:126:

## (i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 466 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

```

5  Met Ala His Leu Pro Gly Gly Ala Ala Ala Pro Leu Ser Glu Asp
   1           5           10           15
Ala Ile Pro Ser Pro Arg Glu Arg Thr Glu Asp Trp Pro Pro Cys Gln
   20           25           30
Ile Val Leu Gln Gly Ala Glu Leu Asn Gly Ile Leu Gln Ala Phe Ala
10  35           40           45
Pro Leu Arg Thr Ser Leu Leu Asp Ser Leu Leu Val Val Gly Asp Arg
   50           55           60
Gly Ile Leu Val His Asn Ala Ile Phe Gly Glu Gln Val Phe Leu Pro
65           70           75           80
15  Leu Asp His Ser Gln Phe Ser Arg Tyr Arg Trp Gly Gly Pro Thr Ala
   85           90           95
Ala Phe Leu Ser Leu Val Asp Gln Lys Arg Ser Leu Leu Ser Val Phe
  100          105          110
Arg Ala Asn Gln Tyr Pro Asp Leu Arg Arg Val Glu Leu Thr Val Thr
20  115          120          125
Gly Gln Ala Pro Phe Arg Thr Leu Val Gln Arg Ile Trp Thr Thr Ala
  130          135          140
Ser Asp Gly Glu Ala Val Glu Leu Ala Ser Glu Thr Leu Met Lys Arg
145          150          155          160
25  Glu Leu Thr Ser Phe Ala Val Leu Leu Pro Gln Gly Asp Pro Asp Val
   165          170          175
Gln Leu Arg Leu Thr Lys Pro Gln Leu Thr Lys Val Val Asn Ala Val
   180          185          190
Gly Asp Glu Thr Ala Lys Pro Thr Thr Phe Glu Leu Gly Pro Asn Gly
30  195          200          205
Lys Phe Ser Val Phe Asn Ala Arg Thr Cys Val Thr Phe Ala Ala Arg
   210          215          220
Glu Glu Gly Ala Ser Ser Ser Thr Ser Ala Gln Val Gln Ile Leu Thr
225          230          235          240
35  Ser Ala Leu Lys Lys Ala Gly Gln Ala Ala Ala Asn Ala Lys Thr Val
   245          250          255
Tyr Gly Glu Asn Thr Thr Phe Ser Val Val Val Asp Asp Cys Ser Met
   260          265          270
Arg Ala Val Leu Arg Arg Leu Gln Val Gly Gly Gly Thr Leu Lys Phe
40  275          280          285
Phe Leu Thr Ala Asp Val Pro Ser Val Cys Val Thr Ala Thr Gly Pro
   290          295          300
Asn Ala Val Ser Ala Val Phe Leu Leu Lys Pro Gln Arg Val Cys Leu

```



305                                      310                                      315                                      320  
 Asn Trp Leu Gly Arg Thr Pro Gly Ser Ser Thr Gly Ser Leu Ala Ser  
    325                                      330                                      335  
 Gln Asp Ser Arg Ala Gly Pro Thr Asp Ser Gln Asp Phe Ser Ser Glu  
 5                                      340                                      345                                      350  
 Pro Asp Ala Gly Asp Arg Gly Ala Pro Glu Glu Glu Gly Leu Glu Gly  
    355                                      360                                      365  
 Gln Ala Arg Val Pro Pro Ala Phe Pro Glu Pro Pro Gly Thr Lys Arg  
    370                                      375                                      380  
 10 Arg His Ala Gly Ala Glu Val Val Pro Ala Asp Asp Ala Thr Lys Arg  
    385                                      390                                      395                                      400  
 Pro Lys Thr Gly Val Pro Ala Ala Pro Thr Arg Ala Glu Ser Pro Pro  
    405                                      410                                      415  
 Leu Ser Ala Arg Tyr Gly Pro Glu Ala Ala Glu Gly Gly Gly Asp Gly  
 15                                      420                                      425                                      430  
 Gly Arg Tyr Ala Cys Tyr Phe Arg Asp Leu Gln Thr Gly Asp Asp Ser  
    435                                      440                                      445  
 Pro Leu Ser Ala Phe Arg Gly Pro Gln Arg Pro Pro Tyr Gly Phe Gly  
    450                                      455                                      460  
 20 Leu Pro  
    465

## (2) INFORMATION FOR SEQ ID NO:127:

## 25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 331 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

30

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

35 Val Cys Pro Pro Pro Pro Thr Asn Met Ala Val Val Cys Gly Ser Gly  
     1                                      5                                      10                                      15  
 Leu Arg Leu Arg Pro Phe His Pro Pro Ser Pro Ser Phe Phe Val Leu  
    20                                      25                                      30  
 Arg Ala Leu Ile Arg Ala Gly Pro Gly Pro Phe Ala Asp Arg Ala Pro  
 40                                      35                                      40                                      45  
 Ser Gly Pro Gly Cys Gly Met Cys Arg Gly Asp Ser Pro Gly Val Ala  
    50                                      55                                      60  
 Gly Gly Ser Gly Glu His Cys Leu Gly Gly Asp Asp Gly Asp Asp Gly  
    378

	65		70		75		80									
	Arg	Pro	Arg	Leu	Ala	Cys	Val	Gly	Ala	Ile	Arg	Phe	Ala	His	Leu	Trp
				85					90						95	
	Leu	Gln	Ala	Thr	Thr	Leu	Gly	Phe	Val	Gly	Ser	Val	Val	Leu	Ser	Arg
5				100					105						110	
	Gly	Pro	Tyr	Ala	Asp	Ala	Met	Ser	Gly	Ala	Phe	Val	Ile	Gly	Ser	Thr
				115					120						125	
	Gly	Leu	Gly	Phe	Leu	Arg	Ala	Pro	Pro	Ala	Phe	Ala	Arg	Pro	Pro	Thr
				130					135						140	
10	Arg	Val	Cys	Ala	Trp	Leu	Arg	Leu	Val	Gly	Gly	Gly	Ala	Ala	Val	Trp
	145					150						155				160
	Ser	Leu	Gly	Glu	Ala	Gly	Ala	Pro	Pro	Gly	Val	Pro	Gly	Pro	Ala	Thr
					165					170					175	
	Gln	Cys	Leu	Ala	Leu	Gly	Ala	Ala	Tyr	Ala	Ala	Leu	Leu	Val	Leu	Ala
15				180					185						190	
	Asp	Asp	Val	His	Pro	Leu	Phe	Leu	Leu	Ala	Pro	Arg	Pro	Leu	Phe	Val
				195					200						205	
	Gly	Thr	Leu	Gly	Val	Val	Val	Gly	Gly	Leu	Thr	Ile	Gly	Gly	Ser	Ala
				210					215						220	
20	Arg	Tyr	Trp	Trp	Ile	Asp	Pro	Arg	Ala	Ala	Ala	Ala	Leu	Thr	Ala	Ala
	225					230					235					240
	Val	Val	Ala	Gly	Leu	Gly	Thr	Thr	Ala	Ala	Gly	Asp	Ser	Phe	Ser	Lys
					245					250					255	
	Ala	Cys	Pro	Arg	His	Arg	Arg	Phe	Cys	Val	Val	Ser	Ala	Val	Glu	Ser
25				260					265						270	
	Pro	Pro	Pro	Arg	Tyr	Ala	Pro	Glu	Asp	Ala	Glu	Arg	Pro	Thr	Asp	His
				275					280						285	
	Gly	Pro	Leu	Leu	Pro	Ser	Thr	His	His	Gln	Arg	Ser	Pro	Arg	Val	Cys
				290					295						300	
30	Gly	Asp	Gly	Ala	Ala	Arg	Pro	Glu	Asn	Ile	Trp	Val	Pro	Val	Val	Thr
	305					310					315					320
	Phe	Ala	Gly	Ala	Leu	Ala	Ala	Cys	Ala	Arg	Ser					
					325					330						

35 (2) INFORMATION FOR SEQ ID NO:128:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2342 base pairs
- (B) TYPE: nucleic acid
- 40 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

	GGCGGGATCT	GCGCACGCGC	GGCACGGEGG	CGGAGAAAGC	GGCGGCAGAG	CCGAAAAGG	60
	CCGGGGGAGG	AAGCGCGGCA	TCCGCGGGGG	GACTCGGTGT	GGGTGGCGAG	GGCCGTGGGT	120
5	CGTCGCGAGG	GGCCACGGGC	ACGCGCCCCG	TGTTTTGTG	AGGCGGGACA	CTCGGTCTGT	180
	TTTCGCGAGC	CGTAGCTGCC	GGCCCGATGG	GCCGCGGTGC	GTACTGGGAC	GTGGGGACGG	240
	ACTGATCGGT	GGCGGGGGGG	GGAAGAAGGG	CCGGGGCCGG	ATTGGGCGTG	GGGCCGCCGG	300
	CGTCGTCGGA	CGCCAGCTCC	TCCAGGCCGT	GGATCCAGGC	CCACATGCGA	GGGGGGACGG	360
	GCTCGCCGGT	GGTGGCGTCG	GTGAGGAGAG	TGGGGGCGAG	GACCCCCGGG	TCCGCCTGCC	420
10	GCGCGGGGGG	GGCAGCGGGG	TCCTCGGGAC	CCGATCCGCC	ATCCCCCCCC	GCAAGGTCCC	480
	GCGGGTCGCG	GGCGGCGGTC	GGGGCAGAGG	GACCTGCCTC	GTCGGCGAGG	GGGCGCTGGT	540
	AAACCGGGTG	TCCCGGGAAC	AGCTCCCCCG	TCAGGAGGGA	GGCGTCGAAG	GGCCGCCCGA	600
	GGATGGCCCG	CGCGAAGAAG	GGGTCCGCGT	CGGCGGCGCT	CGCCGCGAGA	ACGTCCCCCG	660
	CGGTAGCCAC	AAACGGAAGC	TCCTCGGTGG	CCTCGCTGCC	CACAAACCGC	ACGTCAGGGG	720
15	GGCCGGGGGG	CTCCGGGGCT	TCCACAAGA	CCGCGACCGG	GGTCATGGAG	ATGTCCACGA	780
	GGACCAGGCA	CGGGGGCCCG	TCGGCGAGAG	GGCGCTCGGC	GATGAGCGCC	GACAGGCGCG	840
	GGAGCTGCGC	CGCCAGACAC	GCGTTTTCGA	TCGGGTTGAG	ATCGGTGTGG	AGGAGGCCGA	900
	CGGCCACGTC	CTCGATGTCG	GACGACACGA	CGTCGCGCAG	GGCGGCGTCC	GGCCCCCGCG	960
	GGCGCGAGTC	GAAGAGCGTC	AGGCACAGTT	CCAGTTCCGA	CTCGCGGGAG	AAGGCCGTGG	1020
20	TGTTGCGGAG	CGCCACCACG	ACGGGCGCGC	CGAGGAGCAC	CGCGGCCAGA	ACCAGGTCCA	1080
	TGGCCGTAAC	GCGCGCGGCG	GGGGTGCGGT	GGGTGCGGCG	GGCCAGCACG	GCCACGTGCT	1140
	GGCCCCGTTG	TCGGTAGAGG	GCGTGCGGGG	CCTCGGGGAG	GGACGCCTCG	CGCCCCCCCC	1200
	CCGGGCCGAG	CGTCTGGCCA	GACTCCAGGC	GTGCGGCCAG	GAGGGCGTCG	AAGCTGTCTG	1260
	ACTCGGTGTA	GTCGTCGGGA	AACATGCGAG	TCCACAGCGC	GGCCAAAGCG	GCGCTCGGCA	1320
25	GACACATGCG	CCCGAGGACG	CTCACCGCCG	CCAGGGCCTG	GGCCGGACTG	AGCTTCCCGA	1380
	GCGCCGGGGC	GTCCCGGCGC	TGGGTCCCGA	GCTCCAAGGC	CGAGCGCCAG	GGCGCCAGCG	1440
	GGTCGGTTTC	GGACAGCTTG	CCCCGGCGCC	AGTCGGCCAG	CCGCGTGCCG	AACAGGAGGC	1500
	CCCGGGTCGG	GGGGCCTCCG	TCCAAAACG	TCGGCAACAC	GCGGATGCGG	GCGTCGGGAT	1560
	GCGGGGTCAG	GCGCTGGACG	AACAGCATGG	ACTCCGCTGC	GTCTTCGAAC	GCGCGTTCGA	1620
30	GGGTGAGGTG	CATGTACTCG	TGCTGGCGAA	CGAGGTCCAG	GCGCCAGAAG	TTGTAGATGT	1680
	GTTCCGGAAC	GCCGGCCACC	AGCGCGACCA	GCACGTCGTT	CTCGTTGAAG	GCGACGCAGT	1740
	GGCGCTGGGA	CCCCCGGGGG	CCCGGCGGCG	GACGCGGCGC	CGCCGCTCCG	GACGCCACAG	1800
	CCAGCTGGGC	CCAGCGACAC	CCAAACTCGC	GCGTGAGGGT	GGTGGCGACG	AGGGCGACGT	1860
	ACAGCTCGGC	CGCCGCGTCC	ATCGAGGCGC	CCCACGTCGC	CTGGCGATGG	CGCACGAAGC	1920
35	GACCGAACAG	CTGAAAGTTG	GCGGCCTGGG	CGTCGCTGAG	GGCCAGCTGG	AGCCGGTTCA	1980
	CGACGGTCAG	CACGTACATG	GCCGTGACCG	TCGGGGCCGA	TTGAGGACG	TCCGTCGGAA	2040
	GCGGGGGCCG	CACGCAAGCC	GCCTCGGGAC	GCATCAGCAG	CGCGCCGAGT	TTGTGCGTGA	2100
	CGGCCGGGAA	GCATAGCGCG	TACTGCAGCG	GCGTTCGCTC	CGGGGCCAAA	AAGCTGGTGG	2160
	CGAACGGCAG	ATCCAGAGCG	CTGACGGCCT	CACGCAGCAC	CAGGGGCCCC	GGGTCTCCGC	2220
40	CGGCGCGCAG	ATACGCCTCG	CCCCGGCGGC	GCAGCAGCTG	CGGGTCGACC	TCGTGGCCCT	2280
	CGGGGAAGA	AGAGGCCCGG	GCGCGGGCGT	CGAGGGCGCG	AAGATCAACG	AGCAGGGGCG	2340
	C						2342

## (2) INFORMATION FOR SEQ ID NO:129:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 771 amino acids

5 (B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Ala Pro Leu Leu Val Asp Leu Arg Ala Leu Asp Ala Arg Ala Arg Ala  
 1 5 10 15  
 15 Ser Ser Ser Pro Glu Gly His Glu Val Asp Pro Gln Leu Leu Arg Arg  
 20 25 30  
 Arg Gly Glu Ala Tyr Leu Arg Ala Gly Gly Asp Pro Gly Pro Leu Val  
 35 40 45  
 Leu Arg Glu Ala Val Ser Ala Leu Asp Leu Pro Phe Ala Thr Ser Phe  
 20 50 55 60  
 Leu Ala Pro Asp Gly Thr Pro Leu Gln Tyr Ala Leu Cys Phe Pro Ala  
 65 70 75 80  
 Val Thr Asp Lys Leu Gly Ala Leu Leu Met Arg Pro Glu Ala Ala Cys  
 85 90 95  
 25 Val Arg Pro Pro Leu Pro Thr Asp Val Leu Glu Ser Ala Pro Thr Val  
 100 105 110  
 Thr Ala Met Tyr Val Leu Thr Val Val Asn Arg Leu Gln Leu Ala Leu  
 115 120 125  
 Ser Asp Ala Gln Ala Ala Asn Phe Gln Leu Phe Gly Arg Phe Val Arg  
 30 130 135 140  
 His Arg Gln Ala Thr Trp Gly Ala Ser Met Asp Ala Ala Ala Glu Leu  
 145 150 155 160  
 Tyr Val Val Ala Thr Thr Leu Thr Arg Glu Phe Gly Cys Arg Trp Ala  
 165 170 175  
 35 Gln Leu Gly Trp Ala Ser Gly Ala Ala Ala Pro Arg Pro Pro Pro Gly  
 180 185 190  
 Pro Arg Gly Ser Gln Arg His Cys Val Ala Phe Asn Glu Asn Asp Val  
 195 200 205  
 Leu Val Val Ala Gly Val Pro Glu His Ile Tyr Asn Phe Trp Arg Leu  
 40 210 215 220  
 Asp Leu Val Arg Gln His Glu Tyr Met His Leu Thr Leu Glu Arg Ala  
 225 230 235 240  
 Phe Glu Asp Ala Ala Glu Ser Met Leu Phe Val Gln Arg Leu Thr Pro

245 250 255  
 His Pro Asp Ala Arg Ile Arg Val Leu Pro Thr Phe Leu Asp Gly Gly  
 260 265 270  
 Pro Pro Thr Arg Gly Leu Leu Phe Gly Thr Arg Leu Ala Asp Trp Arg  
 5 275 280 285  
 Arg Gly Lys Leu Ser Glu Thr Asp Pro Leu Ala Pro Trp Arg Ser Ala  
 290 295 300  
 Leu Glu Leu Gly Thr Gln Arg Arg Asp Ala Pro Ala Leu Gly Lys Leu  
 305 310 315 320  
 10 Ser Pro Ala Gln Ala Val Ser Val Leu Gly Arg Met Cys Leu Pro  
 325 330 335  
 Ser Ala Ala Ala Leu Trp Thr Cys Met Phe Pro Asp Asp Tyr Thr Glu  
 340 345 350  
 Tyr Asp Ser Phe Asp Ala Leu Leu Ala Ala Arg Leu Glu Ser Gly Gln  
 15 355 360 365  
 Thr Leu Gly Pro Ala Gly Gly Arg Glu Ala Ser Leu Pro Glu Ala Pro  
 370 375 380  
 His Ala Leu Tyr Arg Pro Thr Gly Gln His Val Ala Val Leu Ala Ala  
 385 390 395 400  
 20 Ala Thr Thr Pro Ala Ala Arg Val Thr Ala Met Asp Leu Val Leu Ala  
 405 410 415  
 Ala Val Leu Leu Gly Ala Pro Val Val Val Arg Asn Thr Thr Ala Phe  
 420 425 430  
 Ser Arg Glu Ser Glu Leu Glu Leu Cys Leu Thr Leu Phe Asp Ser Arg  
 25 435 440 445  
 Pro Gly Gly Pro Asp Ala Ala Leu Arg Asp Val Val Ser Ser Asp Ile  
 450 455 460  
 Glu Thr Trp Ala Val Gly Leu Leu His Thr Asp Leu Asn Pro Ile Glu  
 465 470 475 480  
 30 Asn Ala Cys Leu Ala Ala Gln Leu Pro Arg Leu Ser Ala Leu Ile Ala  
 485 490 495  
 Glu Arg Pro Leu Ala Asp Gly Pro Pro Cys Leu Val Leu Val Asp Ile  
 500 505 510  
 Ser Met Thr Pro Val Ala Val Leu Trp Glu Ala Pro Glu Pro Pro Gly  
 35 515 520 525  
 Pro Pro Asp Val Arg Phe Val Gly Ser Glu Ala Thr Glu Glu Leu Pro  
 530 535 540  
 Phe Val Ala Thr Ala Gly Asp Val Leu Ala Ala Ser Ala Ala Asp Ala  
 545 550 555 560  
 40 Asp Pro Phe Phe Ala Arg Ala Ile Leu Gly Arg Pro Phe Asp Ala Ser  
 565 570 575  
 Leu Leu Thr Gly Glu Leu Phe Pro Gly His Pro Val Tyr Gln Arg Pro  
 580 585 590  
 382

Leu Ala Asp Glu Ala Gly Pro Ser Ala Pro Thr Ala Ala Arg Asp Pro  
 595 600 605  
 Arg Asp Leu Ala Gly Gly Asp Gly Gly Ser Gly Pro Glu Asp Pro Ala  
 610 615 620  
 5 Ala Pro Pro Ala Arg Gln Ala Asp Pro Gly Val Leu Ala Pro Thr Leu  
 625 630 635 640  
 Leu Thr Asp Ala Thr Thr Gly Glu Pro Val Pro Pro Arg Met Trp Ala  
 645 650 655  
 Trp Ile His Gly Leu Glu Glu Leu Ala Ser Asp Asp Ala Gly Gly Pro  
 10 660 665 670  
 Thr Pro Asn Pro Ala Pro Ala Leu Leu Pro Pro Pro Ala Thr Asp Gln  
 675 680 685  
 Ser Val Pro Thr Ser Gln Tyr Ala Pro Arg Pro Ile Gly Pro Ala Ala  
 690 695 700  
 15 Thr Ala Arg Glu Trp Ser Val Pro Pro Gln Gln Asn Thr Gly Arg Val  
 705 710 715 720  
 Pro Val Ala Pro Arg Asp Asp Pro Arg Pro Ser Pro Pro Thr Pro Ser  
 725 730 735  
 Pro Pro Ala Asp Ala Ala Leu Pro Pro Pro Ala Phe Ser Gly Ser Ala  
 20 740 745 750  
 Ala Ala Phe Ser Ala Ala Val Pro Arg Val Arg Arg Ser Arg Xaa Xaa  
 755 760 765  
 Xaa Xaa Xaa  
 770  
 25

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14927 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

GCACGTCAGC GCGGCCCCGC GTCGCGATGT CGCCCCAGCT CTCCGGCCCC TGCGCCCCTG 60  
 GCTCGGGGCC GCGCTCCCCG TCCTCGCTCG CGGGCGTCCC CGCGCCACGC CTCCGCCCCC 120  
 CCTCCTCCGC GCGGCCCCGG GGCTCTTCCT CCTCGGCCCC CCCGGTCGCG CCGCCGGCCC 180  
 40 CCAGCCGCGC CAGCACGCGG CGCAGCGCCT CCTCGTCGCA CTGCTCGGGG CTGACGAGCC 240  
 GCCGCAGCAG CGGCGTCGTC AGGTGGTGGT CGTAGCACGC GCGTATCAGC GCCTCGATCT 300  
 GATCGTCGGG CGACGTCGCC TGGCCGCCGA TGATCAGGGC GTCCACCATG TCCAGCGCCG 360  
 CCAGGTGGCC CCCGAACGCG CGATCGAAGT GCTCCGCCCC CCGCCCGAAC AGCGCCAGCT 420

	CCACGGCCAC	CGCGGCGGTC	TCCTGCTGCA	GCTCGCGCTG	CGCCAGCGCG	TTCAGGTTGT	480
	CGGCGAAGGC	GTCCATGGTG	GAGTGGCGGG	CGCGATCGCC	GGACGCCAGC	CAGAAGCGCA	540
	GCTCGCTGAT	GGCGTACAGG	CCGGGCGTAG	TGGCCTGAAA	CACGTCATGC	GCCTCCAGCA	600
	GGGCGTCGGC	CTCCTCGCGG	ACAGAAGAGC	TATCGGCGGG	CGGCGGGCCG	GCCCGGGCCC	660
5	CGCCGCCCGC	CGCGGTCCGC	GCCAGCGCCT	GGTCCAGCAC	ACAGAGCGCT	CGCGCGCGGG	720
	CGGCGTCCGA	CAGCCCGGCG	GCGTGGGGCA	GGTACCGTCG	CAGCTCGTTG	GCGTCCAGCC	780
	GCACCTGGGC	CTGTTGGGTG	ACGTGGTTAC	AGATGCGGTC	CGCCAGGCGG	CGGGCGATGG	840
	TCGCCCCCTG	GTTTCGCGTG	ACGCACAGCT	CCTCGAAACA	GACCGCGCAC	GGGTGGGACG	900
	GGTTCGCTCAG	CTCCGGGGGC	ACGATGAGGC	CCGACCCAC	CGCCGCCACC	ATAAACTCCC	960
10	GGACGCGCTC	CAGCGCGGCC	GTGGCGCCGC	TCGGGGGGGT	GATGAGGTGG	CAGTAGTTCA	1020
	GCTGCTTGAG	AAAATTCTCG	ACATCATGCA	GGAAGCACAG	CTCCATGCGG	ACGTCCCCGC	1080
	CGTACGTCTC	CAGCCGGATC	TGCTGGTGGT	ACGGACAGGG	TCGGGCCAGA	CCCATGGTCT	1140
	CGGTGAAAAA	GGCAGAGACG	TCACCCGTGG	TCGCGAACGT	TTCCAGGTGG	CCCAGGAGCC	1200
	GCTCCCCCTC	GCGCCACGCG	TACTCCAGGA	GCAACTCCAG	GGTGACCGAC	AGCGGGGTGA	1260
15	GAAAGGCGGC	GGCCTGAGCC	TCCAGCCCCG	GCCGCAGGTG	CCGCCGCAGC	ACGCGCACCT	1320
	GGAGCGCGTT	GAGCTTTAGC	TGGGCGAGCT	TCCCCAGGCC	GATCTGGGGG	TCGCATCGTC	1380
	GAAGCAGCTC	TAGCTGAAAA	ACGTACGTCT	GTACCTGCCC	GAGCAGGGCC	AACAGTTTCT	1440
	GTCGGGCCGC	AGTGGGCTCG	GAAACCGCGG	CCGGGGGCGC	GGCCGCCATG	GCGAGTCGCC	1500
	CGGCCGTGCT	GTGGTTTAGT	TAAGGTTTGG	GGGGGTGGGT	CAGAGGCGCG	CCCCGCGCGG	1560
20	ACTGATGCGG	CGGCGGGCCC	CTGACATCCC	CTCTTTATGC	CCGTCGCCCG	CCCGCCCGCC	1620
	CCGCCGGTGT	GCCGTGATTC	GCGGAGTCGG	GGCCTTGTGT	TTCTTTCTTT	CCCCCCCCGA	1680
	ATCCGTTCTT	TCTTCTCAC	CCCCCTCCC	CACACACCCA	CCCAGGACTC	GCCACCACAA	1740
	GGAGGCGAGA	GCCCGTCGCT	AACCCAAAGA	CACAGTCACG	AGACACGATA	TCGACTGTAG	1800
	TTGCGATCGT	TTATTTTATA	CACAACACCA	ACCTTTCCTT	CGACCCCCC	CACCCCCGCC	1860
25	CCTAGAGCAT	ATCCAACGTC	AGGTCTTTT	TCTCCGGTGG	TCCCTCCCCA	AACGGATCGT	1920
	CGCCGTGAAA	CGCCCGCTTT	CGGGCGACGC	CGGCCGCCCC	CGCCGCCGCC	GCCAAACCGC	1980
	CGAACGACGC	CGCGTGGTCA	TCCTCGTCGC	CGAAATCCCC	AAAGTTAAAC	ACCTCCCCGG	2040
	CGGCGCCGAG	CTGGCTGACC	AGGGCCTCCG	CCTCGTGGGC	CACCTCCAGG	GCCGCGTCGG	2100
	TCGACCACTC	GCCGTGCCCG	CGCTCCAGGG	CGCGGGTGGT	AAACTCCATC	ATTTCTCTCG	2160
30	TCAGGTACTC	GTCTCCAGC	AGCGCCAGCC	AGTCCTCGAT	CTGCAGCTGC	TGGGTGCGGG	2220
	GGCCAGGCT	CTTGACGGTC	GCCACAAACA	CGCTGCTGGC	GACCGCCGCC	CCGCCCTCCG	2280
	CAATGATGCC	CCGGAGCTGC	TCGCACAGCG	AATGCTCGTG	GGCCCCGCCC	CCGAGACTCG	2340
	ACGCCGCGCA	CACAAACCCG	GCCCTGGGGC	AGGCCAGGAC	AAACTTGCGG	GTGCGGTCAA	2400
	AGATCAGCAG	CGGGCACGCG	TTTTTGCCGC	CCAGCAGGCT	GGCCAGTTTC	CCGGCTGAA	2460
35	ACACGCGGTC	GTTGCCGGCC	ATGCCGTAGT	ATTTGCTGAT	GCTGAGGCCC	AGCACGACCA	2520
	TCGGGCGCGC	GGCCATCACG	GGCCGCAGCA	GGTTGCAGCT	CGCGAACATG	GACGTCCAGG	2580
	CGCCGGGGTG	CGCGTCGAGG	GAGTCCATCA	GCGCGCGGGC	CCCGGCCCTC	AGGCCCGCGC	2640
	CGCCCTGCGG	GGCCAGGCG	GCGGCCGCTT	GCACGCCGGG	GGGACGGCGG	GACCCGGCGA	2700
	TGACGGCCGT	GAGGGTGTTT	ATGAAGTACG	TCGAGTGGTC	CGAGTACCTC	AAGATCTGGT	2760
40	TGGCCATGTA	GTACATGGCC	AGTTCGCTCA	CGTTATTGGG	GGCCAGGTTC	ATAAAGTTAA	2820
	TCGCGCCGTA	GTCCAGGGAG	AACCTCTTAA	TGAACGCGAT	GGTCTCTATG	TCCTCGCGCG	2880
	ACAAGAGCCG	GGCGGGGAGC	TGGTTGCGCT	GGAGGGCGGT	CCAGAACCAC	TGCGGGTTCG	2940
	GCTGGTTCTGA	CCCCGGGGGC	TTGCCGTTGG	GAAAGATGAC	CGCGTGGAAC	TGCTTCAGCA	3000

	GGAAGCCCAG	CGGTCCGAGG	AGGATGTCCA	CGCGCTTGTC	GGGCTTCTGG	TAGGCGCTCT	3060
	GGAGGCTGGC	GACCCGCGCC	TTGGCGGCCCT	CGGACGCGTT	GGCGCTCGCG	CCCGCGAACA	3120
	ACACGCGGCT	CTTGACGCGC	AGTCCTTTGG	GAAACCCAG	GGTCACGCGG	GCAACGTCGC	3180
	CCTCGAAGCT	GCTCTCGGCG	GGGGCCGTCT	GGCCGGCCGT	TAGGCTGGGG	GCGCAGATAG	3240
5	CCGCCCCCTC	CGAGAGCGCG	ACCGTCAGCG	TCTTCGCCGA	CAGGAACCCG	TTGTTGAACA	3300
	GGTCCATGAC	GCGCCGCCGC	AGCACCGGTT	GGAATTGATT	GCGAAAGTTG	CGCCCCTCGA	3360
	CCGACTGCCC	GGCGAACACC	CCGTGGCACT	GGCTCAGGGC	CAGGTCCTGG	TACACGGCGA	3420
	GGTTGGACCG	CCGCGCGAGG	AGCTGCAGCA	GGGGGCACGG	CCCGCAGGTG	TACGGGTCCA	3480
	GCGACAGCGA	CATGGCGTGG	TTGGCCTCGG	CCAGACCGTC	GCGGAACTTA	AAGTTGCGCC	3540
10	CCTCGATCAG	GTTGCGCATC	AGCTGTTCCA	CCTCGCGATC	CACCAGCTGC	TTGATGTTGT	3600
	TCACCACCGT	GTGCAGGGCC	TCGCGGGTGC	CGATAATCGT	CTCCAGCCTC	CCCAGGGCCG	3660
	TGGGCACCGC	CTGGTCCACG	TACTGCAGGG	CCTCGAGCTC	GGCCATGACG	CGCTCGGTGG	3720
	CCGCGCGGTA	CGTCTCCTGC	ATGATGGTCC	GGGTGTTCTC	GGACCCGTCC	GCGCGCTTCA	3780
	GGGCCGAGAA	GGCGGCGTAG	TCCCCAGCA	CGTCGCAGTC	GCTGTACGCG	CTGTTTCATCG	3840
15	TTCCGAAGAC	CCCAATGGCC	CCCCGGGCGG	CGCTCGCGAA	CTTGGGGTGG	CGGGCCCCGA	3900
	GCCGCATCAG	CGTCGTGTGC	GCGCAGGCGT	GGCGGGTCTC	GAAGGTACAC	AGGTTGCAGG	3960
	GCACGTCGGT	CTGGCCCGAG	TCCGCGACGT	AGCGAAACAC	GTCCATCTCC	TGGCGCCCGA	4020
	CGATGACTCC	GCCGTCGCAG	CGCTCCAGGT	AAAACAGCAT	CTTGGCCAGC	AGGGCCGGAG	4080
	AGAACCCGCA	CAGCATGGCC	AGGTGCTCGC	CGGCGAATC	CTGGGTTCGG	CCGACGAGGG	4140
20	GCGCCGTGGG	GCGCCCCCTG	TACCCGGGCA	CCACGTGGCC	CTCGCGGTCC	AGCTGCGGGT	4200
	TGGCCGCCAC	GTGCGTGCCG	GGCAGGAGAA	AGAAGCGGTA	AAAGGAGGGC	TTGCTGTGGT	4260
	CCTTGGGGTC	CGCCGGCCCC	GCGTCGTCCA	CCTCGGTCAG	GTGGAGGGCC	GAGTTGGTGC	4320
	TGAACACCAT	GGCGCCACG	AGGCCCGCGG	CGCGCGCCAG	GTACGCCCCG	ACGGCGCCCG	4380
	CACGGGCCGC	GGGCGTTTCC	TGGCCCTCAA	GCAGGGGCCA	CGTGGTGATG	TCGGGGGGCG	4440
25	GCTCGTCAAA	GACCGCCATC	GACACGATGG	ACTCCAGGGC	CAGGGCGGCG	TCGCCCGCCA	4500
	TCACCGAGGC	CAGGCGCTGC	TCAAACCCGC	CCGCCGGGCC	CTTGTTCCCG	GCGTCGCGCG	4560
	CGCCCCGCTG	GGGCTTACCC	TGGCTGGCCT	CGAAGGCCGT	GAACGTAATG	TCGGCGGGGA	4620
	GGGCCGCGCC	CTCGTGGTTT	TCGTCGAACG	CCAGGTGGGC	GGCCGCGCGG	GCCACGGCGT	4680
	CCACGTTCCG	AGCACGCGAG	GCCACGGCGG	CGGGCCCGAC	GACCGCCTCG	AACAGCAGGC	4740
30	GGGCGAGGGG	GCGGTTGAAA	AACGGAAGGG	GGTAGTTGAA	ATTCTCCCCG	ATCGATCGGT	4800
	GGTTGCAGTT	AAACGGATCG	GCGATGACCC	GGCTAAAATC	CGGCATAAAC	ATCTGCAGCG	4860
	GATACACGGG	GATGCGGTGA	ACCTCCGCGT	CCCCGATGGT	TACCTTGTTCC	ATCCCGCCCA	4920
	GGTGCAGGAA	GGTGTGCTG	ATGCACACGG	CCTCCCGGAA	GCCCTCCGTG	ATCACCAGAT	4980
	ACAGCAAGGC	CCGGTCCGGG	TCCAGTCCGA	GCCGCTCGCA	CAGCGCGTCC	CCCGTCGTCT	5040
35	CGTGCTTTAG	GTCGCGAGGG	CGGGGCGCGT	AGTCCGAGAA	GCCAAAATGG	CGGCGCGCCC	5100
	GCTCGCAGAG	CCGCGTCAGG	TTGGGGGCCCT	GGGTGCTGGG	GGCCAGGTGG	CGGCCGCCGT	5160
	GAAAGACGTA	GACGGACGGG	CTGTAGTGCG	AGGGCATAAG	CTTGAGGGAC	ACCGCGGTCC	5220
	CCCCAAGGCC	CGTCGTGCGG	GACCCGACGA	CCGCGGCCAC	GTTGGCCTCA	AACCCGCTCT	5280
	CCACGGTCAG	GCCGACGATG	AGGGGCGCGA	CGGCGACGTC	CGCGTCGCCG	CTGCGCGCCG	5340
40	ACAGTAGCGA	CAGCAGCTCC	AGGCCTTCGG	CCGACAGGC	GCGGCCATAC	ACGTACCCCA	5400
	TCGGCCCCGG	AGGAACCTTG	ACGGTGGTCG	TCGTTTGGG	CTTGGTGTCC	ATGGCTTTTCG	5460
	GGAGATTGGC	GACCGGCAGG	AACGGGGGCC	CGGCAAGACG	ACCGGGGGCA	GACGGGGGAG	5520
	GCCGCGCGTG	GTCGACGGCT	GCTGCCCGCC	GTCGTCTCTC	CGATGGGGTC	GAATGCCGGC	5580



	GCTGGGGGTG	GGGTCTACAC	CCGCCCCTTC	ACCGAGCGGC	CCCTGGTGGG	GGTGGGATGG	5640
	GTGGGATGGG	GTGGGCGAGA	ATGGCCCCGC	ACCGGATCGC	GCCGGACGGG	GGGGCCCCGG	5700
	GTGGGGCAAG	GTTTGGGCGC	AAGGCTCCAG	CGGCGATTTC	AGAGGCCTGC	GGATGGCGGC	5760
	CCAGAGCTGG	GTATGCTCGG	CCGGGGCGGC	CGGTATATGT	ACGGCGTGCT	GGGAGGGGCG	5820
5	GCGTCGGGCC	CCGCCCCACG	TCCGCCACGC	CCCGCGCGTC	ATCGGCAGGG	GGCGTGGCCG	5880
	CCCTTCTAAA	AAAAGTGAGA	ACGCGAAGCG	TTCGCACTTT	GTCCTAATAG	TATATATATT	5940
	ATTAGGACAA	AGTGCGAACG	CTTCGCGTTC	TCACTTTTTT	TAGAAGGGCG	GCCACGCCCC	6000
	CTTTGACGTC	ACGCTCACCC	GGGCGGCCGG	CCGCCCATAA	GCGCGGCTTG	CCGGGCCGAT	6060
	AAAAAGAAAC	CGCGGCGCCC	CCGCGGACAC	CACACACTGG	CTCTCGAACC	CCGGACGCGC	6120
10	AGAAGGGACC	CGGGCGCGGG	TCCGCCGGTA	AGAGCCGGGG	GGAACATCGG	CACCGCCATC	6180
	CCACCCCGAG	CTGTTGGGTG	GGCGGGTGGG	GGGGCTGGTG	AGGCGGTGGT	GGGAGGGGGC	6240
	GGCGTATAGC	AGGACAACGA	CCGGCGGCGA	TGTTTTGTGC	CGCGGGCGGC	CCGACTTCCC	6300
	CCGGGGGGAA	GTCGGCGGCT	CGGGCGGCGT	CTGGGTTTTT	TGCCCCCAC	AACCCCGGG	6360
	GAGCCACCCA	GACGGCACCG	CCGCCTTGCC	GCCGGCAGAA	CTTCTACAAC	CCCCACCTCG	6420
15	CTCAGACCGG	AACGCAGCCA	AAGCCCCCG	GGCCGGCTCA	GCGCCATACG	TACTACAGCG	6480
	AGTGCGACGA	ATTTGATTT	ATCGCCCCG	GTTCGCTGGA	CGAGGACGCC	CCCGCGGAGC	6540
	AGCGCACCGG	GGTCCACGAC	GGCCGCTCC	GCGCGCCCC	TAAGGTGTAC	TGCGGGGGGG	6600
	ACGAGCGCGA	CGTCTCCGC	GTGGGCCCGG	AGGGCTTCTG	GCCGCGTCGC	TTGCGCCTGT	6660
	GGGGCGGTGC	GGACCATGCC	CCCGAGGGGT	TCGACCCAC	CGTCACCGTC	TTCCACGTGT	6720
20	ACGACATCCT	GGAGCACGTG	GAACACGCGT	ACAGCATGCG	CGCCGCCAG	CTCCACGAGC	6780
	GATTTATGGA	CGCCATCACG	CCCGCCGGGA	CCGTCATCAC	GCTTCTGGGT	CTGACCCCG	6840
	AAGGCCATCG	CGTCGCCGTT	CACGTCTACG	GCACGCGGCA	GTACTTTTAC	ATGAACAAGG	6900
	CAGAGGTGGA	TCGGCACCTG	CAGTGCCGTG	CCCCGCGCGA	TCTCTGCGAG	CGCCTGGCGG	6960
	CGGCCCTGCG	CGAGTCGCCG	GGGGCGTCGT	TCCGCGGCAT	CTCCGCGGAC	CACTTCGAGG	7020
25	CGGAGGTGGT	GGAGCGCGCC	GACGTGTACT	ATTACGAAAC	GCGCCCGACC	CTGTACTACC	7080
	GCGTCTTCGT	GCGAAGCGGG	CGCGCGCTGG	CCTACCTGTG	CGACAACCTT	TGCCCCGCGA	7140
	TCAGGAAGTA	CGAGGGGGGC	GTCGACGCCA	CCACCCGGTT	TATCCTGGAC	AACCCGGGGT	7200
	TTGTACCTTT	CGGCTGGTAC	CGCCTCAAGC	CCGGCCGCGG	GAACGCGCCG	GCCCAACCGC	7260
	GCCCCCGGAC	GGCGTTCGGA	ACCTCGAGCG	ACGTCGAGTT	TAACGTCACG	GCGGACAACC	7320
30	TGGCCGTGCA	GGGGGCCATG	TGTGACCTGC	CGGCCTACAA	GCTCATGTGC	TTGATATATCG	7380
	AATGCAAGGC	CGGGGGGGAG	GACGAGCTGG	CCTTTCCGGT	CGCGGAACGC	CCGGAAGACC	7440
	TCGTCATCCA	GATCTCCTGT	CTGCTCTACG	ACCTGTCCAC	CACCGCCCTC	GAGCACATCC	7500
	TCCTGTTTTT	GCTCGGATCC	TGCGACCTCC	CCGAGTCCCA	CCTCAGCGAT	CTCGCCTCCA	7560
	GGGGCCTGCC	GGCCCCCGTC	GTCTTGGAGT	TTGACAGCGA	ATTCGAGATG	CTGCTGGCCT	7620
35	TCATGACCTT	CGTCAAGCAG	TACGGCCCCG	AGTTCTGTGAC	CGGGTACAAC	ATCATCAACT	7680
	TCGACTGGCC	CTTCGTCTTG	ACCAAGCTGA	CGGAGATCTA	CAAGGTCCCG	CTCGACGGGT	7740
	ACGGGCGCAT	GAACGGCCGG	GGTGTGTTCC	GCGTGTGGGA	CATCGGCCAG	AGCCACTTTC	7800
	AGAAGCGCAG	CAAGATCAAG	GTGAACGGGA	TGGTGAACAT	CGACATGTAC	GGCATCATCA	7860
	CCGACAAGGT	CAAACCTCTC	AGCTACAAGC	TGAACGCCGT	CGCCGAGGCC	GTCTTGAAGG	7920
40	ACAAGAAGAA	GGATCTGAGC	TACCGCGACA	TCCCCGCCTA	CTACGCCTCC	GGGGCCGCGC	7980
	AGCGCGGGGT	GATCGGCGAG	TATTGTGTGC	AGGACTCGCT	GCTGGTCGGG	CAGCTGTTCT	8040
	TCAAGTTTCT	GCCGCACCTG	GAGCTTTCGG	CCGTCGCGCG	CCTGGCGGGC	ATCAACATCA	8100
	CCCGCACCAT	CTACGACGGC	CAGCAGATCC	GCGTCTTCAC	GTGCCTCCTG	CGCCTTGCGG	8160

	GCCAGAAGGG	CTTCATCCTG	CCGGACACCC	AGGGGCGGTT	TCGGGGCCTC	GACAAGGAGG	8220
	CGCCCAAGCG	CCCGGCCGTG	CCTCGGGGGG	AAGGGGAGCG	GCCGGGGGAC	GGGAACGGGG	8280
	ACGAGGATAA	GGACGACGAC	GAGGACGGGG	ACGAGGACGG	GGACGAGCGC	GAGGAGGTCG	8340
	CGCGCGAGAC	CGGGGGCCGG	CACGTGCGGT	ACCAGGGGGC	CCGGGTCTCT	GACCCACCT	8400
5	CCGGGTTTCA	CGTCGACCCC	GTGGTGGTGT	TTGACTTTGC	CAGCCTGTAC	CCCAGCATCA	8460
	TCCAGGCCCA	CAACCTGTGC	TTCAGTACGC	TCTCCCTGCG	GCCCGAGGCC	GTGCGCACC	8520
	TGGAGGCGGA	CCGGGACTAC	CTGGAGATCG	AGGTGGGGGG	CCGACGGCTG	TTCTTCGTGA	8580
	AGGCCCACGT	ACGCGAGAGC	CTGCTGAGCA	TCCTGCTGCG	CGACTGGCTG	GCCATGCGAA	8640
	AGCAGATCCG	CTCGCGGATC	CCCCAGAGCA	CCCCGAGGA	GGCCGTCTCT	CTCGACAAGC	8700
10	AACAGGCCCG	CATCAAGGTG	GTGTGCAACT	CGGTGTACGG	GTTACCGGG	GTGCAGCACG	8760
	GTCTTCTGCC	CTGCCGAC	GTGGCCGCCA	CCGTGACGAC	CATCGGCCGC	GAGATGCTCC	8820
	TCGCGACGCG	CGCGTACGTG	CACGCGCGCT	GGGCGGAGTT	CGATCAGCTG	CTGGCCGACT	8880
	TTCCGGAGGC	GGCCGGCATG	CGCGCCCCCG	GTCCGTACTC	CATGCGCATC	ATCTACGGGG	8940
	ACACGGACTC	CATTTTCGTT	TTGTGCCGCG	GCCTCACGGC	CGCGGGCCTG	GTGGCCATGG	9000
15	GCGACAAGAT	GGCGAGCCAC	ATCTCGCGCG	CGCTGTTTCT	CCCCCGATC	AAGCTCGAGT	9060
	GCGAAAAAAC	GTTACCAAG	CTGCTGCTCA	TCGCCAAGAA	AAAGTACATC	GGCGTCATCT	9120
	GCGGGGGCAA	GATGCTCATC	AAGGGCGTGG	ATCTGGTGCG	CAAAAACAAC	TGCGCGTTTA	9180
	TCAACCGCAC	CTCCAGGGCC	CTGGTCGACC	TGCTGTTTTA	CGACGATACC	GTATCCGGAG	9240
	CGGCCCGCCG	GTTAGCCGAG	CGCCCCGAG	AGGAGTGGCT	GGCGCGACCC	CTGCCCGAGG	9300
20	GACTGCAGGC	GTTCCGGGCC	GTCTCTGTAG	ACGCCCATCG	GCGCATCACC	GACCCGAGGA	9360
	GGGACATCCA	GGACTTTGTC	CTCACCGCCG	AACTGAGCAG	ACACCCGCGC	GCGTACACCA	9420
	ACAAGCGCCT	GGCCACCTG	ACGGTGTATT	ACAAGCTCAT	GGCCCGCCGC	GCGCAGGTCC	9480
	CGTCCATCAA	GGACCGGATC	CCGTACGTGA	TCGTGGCCCA	GACCCGCGAG	GTAGAGGAGA	9540
	CGGTGCGCGC	GCTGGCCGCC	CTCCGCGAGC	TAGACGCCGC	CGCCCCAGGG	GACGAGCCCC	9600
25	CCCCCCCAGC	GGCCCTGCCC	TCCCCGGCCA	AGCGCCCCCG	GGAGACGCCG	TCGCATGCCG	9660
	ACCCCCCGGG	AGGCGCGTCC	AAGCCCCGCA	AGCTGCTGGT	GTCCGAGCTG	GCGGAGGATC	9720
	CCGGGTACGC	CATCGCCCCG	GGCGTTCCGC	TCAACACGGA	CTATTACTTC	TCGCACCTGC	9780
	TGGGGGCGGC	CTGCGTGACG	TTCAAGGCC	TGTTTGGAAA	TAACGCCAAG	ATCACCGAGA	9840
	GTCTGTTAAA	GAGGTTTATT	CCCAGACGCT	GGCACCCCCC	GGACGACGTG	GCCGCGCGGC	9900
30	TCAGGGCCCG	GGGGTTCGGG	CCGCGGGGGG	CCGGCGCTAC	GGCGGAGGAA	ACTCGTCGAA	9960
	TGTTGCATAG	AGCCTTTGAT	ACTCTAGCAT	GAGCCCCCCG	TCGAAGCTGA	TGTCCCGCAT	10020
	CTTGCAATAA	ATGTCTGCGG	CCGACACGGT	CGGAATTTCC	GCGTCCGCTG	GTTTCTCTGC	10080
	GTTGCGTCTG	ACCACGAGCA	CAAACGTGCT	CTGCCACACG	TGGGCGGCGA	ACCGGTAGCC	10140
	GGGGCACGCG	GTCAGCATCC	GATCGATGAG	CCGGTAGTGC	AGGTGGGCCG	ACGTGCCGGG	10200
35	GAAGATGACG	TACAGCATGT	GGCCCCCGTA	CGTGGGGTCC	GGGTAAAAAA	GAAACCGGGG	10260
	GTGCGACGCC	CCCCCTCCGC	GCAGGATCGT	GTGCACGAAA	AAGAGCTCGG	GCTGGCCGAG	10320
	CGTATCGGCC	AGGAGGTCCT	GGAGGGGGGT	GCTGTGGCGG	TCGGCCAGCA	CGACCAGGGA	10380
	GGCCAGAAAAG	GTGCGGTGCT	CAAAGATCGT	ATTGATCTGC	TGCACGAAGG	CCAGGATGAG	10440
	GGCCTCGCGG	CTGACGGTGG	CCAGCCGCCC	GTCGCCCGCG	CTGCACGCGG	GGCAGCAGCC	10500
40	CCCATCCCC	AGGTAGTAGC	CCATGCCCGA	GAGGGTCAGG	CAGTTGTCGG	CCACGGTCTG	10560
	GTCCAGGCTG	AAGGGGAGCG	ACACGGGGGT	CGTCTTCACC	AGGGGCACGG	ATAGCGAGCG	10620
	CACGATGGCG	ATCTCCTCGG	AGGGCGTCTG	GGCGAGGGCG	GCGAAGAAGC	CGCGGTAGCG	10680
	ACGGCGCTCG	TGCAGGCAGA	GCTCCAGCCT	GCGCGCGTGC	GACGGCAGGC	TCTTGCGGGA	10740

	GGCCCGGCGC	TCCACGCCGG	GGTCCCCGGC	GGCGAAAAAG	CGCGACCGCC	GCCGGGTCTT	10800
	GTCGCGGCCG	GGCCCGGGCC	GGGAGCCGGA	GCGACGGGGG	GCGATGTCAT	ACATAGGTAC	10860
	AGAGGGTGTG	CTCCAGGGAC	AGGAGAGAGA	TCGAGTGTCT	TCTGAGCAGC	GCGCCGGCCT	10920
	CGCGGACAAA	TGTGGCCAGC	GCGGTGGGCT	TCGGCACAAA	TACCTGGTAC	GTCTTGAAGG	10980
5	TGTAGATGAG	GGCCCGCAGG	GCTATACAGA	CCCGCCCCCTC	GAATCGTTG	CCGCAGGCCA	11040
	ACTTGGCCTT	GTGAAGCTGC	AGCTCGTCTC	GATGGTCGGC	GCGGGGGTGG	CCAAACAGGA	11100
	CCCAGGGGTC	GACTTCCATC	TCCGTGATGG	CGCACATGGG	ATCGCAGAAC	ATGTGCTTGA	11160
	AGATGGCCTC	GGGGCCCGCG	GCCCCAAGCA	GGCTCACGAA	CCGGCCCCCG	TCCCCGGGCT	11220
	GCGCCTCGGG	GTCCGCCTCG	AGCTGGTCCA	CGACCGGCAC	TATGCAGTCG	AAGAGGCTGG	11280
10	TGTTGTTCTC	CGAGTAGCGG	ACGACGGACG	CCCTCAGGCG	TCGCATGGCC	AGCCAGTAGG	11340
	CCCGCACCAG	CAACAGATTG	CACAGCAGGC	ATCCCCCGCC	GGTGCGCCCG	CGGCCCCGGC	11400
	CGTGCTTCAG	CACGGTGGCC	ATCAGCGGGC	CCAGGTCCAG	GTCGGGCTGG	GCCTTGGGCT	11460
	CGGCGAAGTC	CGCAAAGCGC	GGGGCCCGCT	CGCGCATGCG	CGCCCCGCGG	TGCGCTTCCC	11520
	AGGACTCGCT	GACCGCGGCG	CGGCGGGCGT	CCGCGGCGGC	GCGCAGCCGG	GGCCCCGACT	11580
15	CCCAGACGGC	GGGGGTGCCG	GCGAGCAGCA	GCAGGATCAG	GTCGGCGTAC	GCCCACGTCT	11640
	CCGGCTCACC	CCCCTGCGCC	AGCGCCCCGG	CGCGCGCCTC	GAACTCCCCG	TTGCGGGCGG	11700
	CGGCGCGCGT	GCAGCAGCTG	TCTCCGCCCC	CGCGCTTGCC	CTCGGTGCAG	TCGAGCAGGC	11760
	GGGCGCAGTC	CTTCCAGTTC	ATCAGGGCGG	TGGTGAGGGA	GGGTGCGGTT	CCCGAGCCCC	11820
	CGCCCGCCCC	CGCCCGTCA	TCGCCCCCGG	AGGCCAGGGT	CCCAGTGAGG	GCCCGGGTTG	11880
20	CGGACTGCGC	GAGGAAGGAA	TAGTTGGAGT	ACTGCACCTT	GGCGGCGCCC	GGGGAGGGCG	11940
	TCGGCCTGGG	TTGCTTCTGG	GCGTGCGGCC	CGGGCACCCC	GCCGTGCGTC	CGGAAGCAGC	12000
	AGTGAGAGAA	GAAATGCCGG	TGGATGTCGT	TGATGGTCAG	GGCGAAGCGC	GCGAAGGAGC	12060
	CGACAAGGGT	CGCCTTCTTG	GTGCGCAGGA	AGTGGTGGTC	CATGACGTAG	ACGAAGTCGA	12120
	AGGCGGCCAC	GAAGATGCTC	GCGGCGCAGT	GGGGCGCGCC	CAGGCACTTG	GCGCAGAGGA	12180
25	ACGCGTAATC	GGCCACCCAC	TGGGGCGAGA	GGCGGTAGGC	CTGCTTGATC	AGCTCGATGG	12240
	TGCGGCAGAC	CAGACAGGGG	CGGTCCAGCG	CGAAGGTGTC	GACGGACGCC	GCGGCGAAGG	12300
	GCCCCGTGTC	CAAGAGTCCC	TCTGCCGTGG	GGTCTGCGGG	GGGGCCGCGG	GCGGACCCCC	12360
	GCCCCCGCCC	CCCCGAAGCC	TCGCGCGCGG	CCCCGCGCGG	CCGCGGGGGG	GCGGGCGCGA	12420
	CGTCGCTCTC	CACGTCTCTG	TCGAGCGCGC	TCGCGGGCGG	CACGCCATCC	ACGTGACAGG	12480
30	CCGCCAGGAG	CTCGGCGCAC	AGGGCCTCGT	TAAGAGCCAG	AAGGTCGGGA	TCGAAGGCCA	12540
	CATACGGACG	CTCGAACGCG	CCCTCCTTCC	AGCTGCTGCC	CGGCGACTCT	TCGCGCACGG	12600
	CGGCGCTCGA	CGGCACCCCC	GGGGCGGACG	TCGCCATGGC	CGGTGAGCGC	GGGCGCACGC	12660
	GTCCGCGAAC	GTTACGGGAC	GCGATCCCCG	ACTGCGCGCT	GCGGTCCCAG	ACCTTGAAAA	12720
	GTCTAGACGC	GCGCTACGTC	TCGCGAGACG	GCGCGGGGGA	CGCGGCCGTC	TGGTTCGAGG	12780
35	ACATGACCCC	CGCCGAACCTA	GAGGTTATAT	TCCCGACCAC	GGACGCCAAG	CTGAAGTACC	12840
	TCTCGCGGAC	GCAGCGGCTG	GCCTCCCTCC	TGACGTACGC	CGGGCCTATA	AAAGCGCCCC	12900
	ACGGCCCCGC	CGCCCCACAT	ACGCAGGACA	CCGCGTGGCT	GCACGGCGAG	CTGCTCGCCC	12960
	GAAAGCGCGA	ACGGTTCGCG	GCGGTCATTA	ACCGGTTCCCT	GGACCTGCAC	CAGATCCTGC	13020
	GGGGCTGACG	CGCGCTTCGG	CGGGGCACCG	GCACCGGGAC	CGACTTGTTT	TACATAACAG	13080
40	TAGGGGGTGG	GGGAACGCGC	ACCCTTGCCC	GGTCGCGATG	GCGGGGATGG	GGAAGCCCTA	13140
	CGGCGGCCGC	CGGGGGGACG	CGTTCGAGGG	TCTCGTTCAG	CGCATCAGGC	TCATTTGTTCC	13200
	CACCACGCTG	CGCGGCGGGG	GTGGGGAGTC	GGGCCCCCTAC	TCGCCATCCA	ACCCGCCCTC	13260
	GAGATGTGCC	TTCCAGTTCC	ACGGCCAGGA	TGGGTCCGAC	GAGGCCCTCC	CGATCGAGTA	13320

CGTCTGCGG CTCATGAACG ACTGGGCCGA TGTGCCCTGC AACCCCTACC TGC GCGTGCA 13380  
 GAACACCGGC GTTTCGGTGC TGTTCAGGG GTTTTAAAC CGGCCCCACG GCGCCCCGGG 13440  
 GGGCGCGATC ACGGCGGAGC AGACCAACGT GATTCTGCAC TCCACCGAGA CGACGGGACT 13500  
 GTCCCTCGGA GACCTGGACG ACGTCAAGGG GCGCCTCGGC CTGGACGCCC GGCCGATGAT 13560  
 5 GGCCAGCATG TGGATCAGCT GCTTTGTGCG CATGCCCCGG GTGCAGCTCG CGTTTCGGTT 13620  
 CATGGGCCCC GAGGACGCCG TTCGCACGCG GCGGATCCTG TGTCGCGCCG CCGAGCAGGC 13680  
 CCTCGCCCGT CGCGCCCGGT CCAGGCGGTC CCAGGATGAC TACGGGGCGG TGGCGGTGGC 13740  
 GGCGGCGCAC CACTCTTCCG GAGCGCCCGG GCCGGGGGTC GCCGCCTCGG GCCCGCCAGC 13800  
 GCCGCCCGGA CGGGGACCGG CCCGTCCGTG GCATCAGGCC GTGCAGTTGT TCCGGGCCCC 13860  
 10 GCGTCCGGGC CCCCCGGCGC TTCTGTTGCT GGTGGCGGGG CTGTTTCTGG GGGCCGTAT 13920  
 CTGGTGGGCG GTTGGCGCGC GCCTATGAAA GGGGCGGAGC CACCGTCCCG CCCGCCAGTG 13980  
 CATCCCAGAC GCCCGCGAGC CGCACATCCC CTCGCTCCC GCCTCCGGCC CGATTCTTAC 14040  
 GGCGGACCC AAGGTCCCGA TGGCCGCCCC GCAGTTTCAC CGCCCCAGCA CCATTACCGC 14100  
 CGACAACGTC CGGGCGCTCG GCATGCGCGG GCTCGTGTG GCCACCAACA ACGCTCAGTT 14160  
 15 CATCATGGAT AACAGCTACC CGCATCCGCA CGGAACGCAG GGTGCGGTGC GAGAGTTTCT 14220  
 TCGCGGGCAG GCCGCGGCGC TGACGGACCT CGGGGTGACC CACGCCAACA ACACGTTCCG 14280  
 CCCGCAGCCT ATGTTGCGCG GCGACGCCGC GGCCGAATGG CTGCGGCCCT CGTTCGGTCT 14340  
 TAAGCGCACG TATTCCTTCT TTGTCGTTCT CGACCCCAAG ACCCCAGCA CCCCGTGAGT 14400  
 CCTCGGCGGG TCCCTCCGCG GCCGTCTCTC GTTGCCCCCT CTTTCCCCCT TCCCGGTGG 14460  
 20 TTCAATAAAA AACACCAACA TACGATATTC GCGTTTGATA CGTTTATTGG GGGGGTGTAG 14520  
 GGCCCAACGA TCGGCGATTA ACAACACCAA ACAATCGAGC GCGTCTAACC CAGTAACATG 14580  
 CGCACGTGAT GTAGGCTGGT CAGCACGGCG TTGCTGCGCT GAAACAGCGC CCTGCGGGTC 14640  
 CGCTGCAGCT GTTGTGTAT GCGGCGGCAT GCGCGGATCA AAACCGCCAG GGCGCTACGA 14700  
 CCGGTGCTTC GTACGTAGCG TCGCGACAAG ACGGCATTTC CCTGTACGGG CAAGGGGCCA 14760  
 25 AATTGCGAGT GTGGTGACTG GAGGTGGTCG GCGGCCAATG GGCCGGGTGG TTCGTCGGCG 14820  
 GGGGCAAGT GCGGTTCCGG TGGGAGGGGG TCGAGCGCCT CGGTATCATC CGAGTCCGAG 14880  
 AAACGCAGGG AGTCTGCGTC GGAGTGTTC TATCGGAGG AGATGT 14927

## (2) INFORMATION FOR SEQ ID NO:131:

30

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 495 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

35

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

40

Met Ala Ala Ala Pro Pro Ala Ala Val Ser Glu Pro Thr Ala Ala Arg  
 1 5 10 15  
 Gln Lys Leu Leu Ala Leu Leu Gly Gln Val Gln Thr Tyr Val Phe Gln

	20	25	30
	Leu Glu Leu Leu Arg Arg Cys Asp Pro Gln Ile Gly Leu Gly Lys Leu		
	35	40	45
5	Ala Gln Leu Lys Leu Asn Ala Leu Gln Val Arg Val Leu Arg Arg His		
	50	55	60
	Leu Arg Pro Gly Leu Glu Ala Gln Ala Ala Phe Leu Thr Pro Leu		
	65	70	75
	Ser Val Thr Leu Glu Leu Leu Leu Glu Tyr Ala Trp Arg Glu Gly Glu		
	85	90	95
10	Arg Leu Leu Gly His Leu Glu Thr Phe Ala Thr Thr Gly Asp Val Ser		
	100	105	110
	Ala Phe Phe Thr Glu Thr Met Gly Leu Ala Arg Pro Cys Pro Tyr His		
	115	120	125
	Gln Gln Ile Arg Leu Glu Thr Tyr Gly Gly Asp Val Arg Met Glu Leu		
15	130	135	140
	Cys Phe Leu His Asp Val Glu Asn Phe Leu Lys Gln Leu Asn Tyr Cys		
	145	150	155
	His Leu Ile Thr Pro Pro Ser Gly Ala Thr Ala Ala Leu Glu Arg Val		
	165	170	175
20	Arg Glu Phe Met Val Ala Ala Val Gly Ser Gly Leu Ile Val Pro Pro		
	180	185	190
	Glu Leu Ser Asp Pro Ser His Pro Cys Ala Val Cys Phe Glu Glu Leu		
	195	200	205
	Cys Val Thr Ala Asn Gln Gly Ala Thr Ile Ala Arg Arg Leu Ala Asp		
25	210	215	220
	Arg Ile Cys Asn His Val Thr Gln Gln Ala Gln Val Arg Leu Asp Ala		
	225	230	235
	Asn Glu Leu Arg Arg Tyr Leu Pro His Ala Ala Gly Leu Ser Asp Ala		
	245	250	255
30	Ala Arg Ala Arg Ala Leu Cys Val Leu Asp Gln Ala Arg Thr Ala Ala		
	260	265	270
	Gly Gly Gly Ala Arg Ala Gly Pro Pro Pro Ala Asp Ser Ser Ser Val		
	275	280	285
	Arg Glu Glu Ala Asp Ala Leu Leu Glu Ala His Asp Val Phe Gln Ala		
35	290	295	300
	Thr Thr Pro Gly Ala Ile Ser Glu Leu Arg Phe Trp Leu Ala Ser Gly		
	305	310	315
	Asp Arg Ala Arg His Ser Thr Met Asp Ala Phe Ala Asp Asn Leu Asn		
	325	330	335
40	Ala Gln Arg Glu Leu Gln Gln Glu Thr Ala Ala Val Ala Val Glu Leu		
	340	345	350
	Ala Leu Phe Gly Arg Arg Ala Glu His Phe Asp Arg Ala Phe Gly Gly		
	355	360	365

His Leu Ala Ala Leu Asp Met Val Asp Ala Leu Ile Ile Gly Gly Gln  
 370 375 380  
 Ala Thr Ser Pro Asp Asp Gln Ile Glu Ala Leu Ile Arg Ala Cys Tyr  
 385 390 395 400  
 5 Asp His His Leu Thr Thr Pro Leu Leu Arg Arg Leu Val Ser Pro Glu  
 405 410 415  
 Gln Cys Asp Glu Glu Ala Leu Arg Arg Val Leu Ala Arg Leu Gly Ala  
 420 425 430  
 Gly Gly Ala Thr Gly Gly Ala Glu Glu Glu Glu Pro Arg Ala Ala Ala  
 10 435 440 445  
 Glu Glu Gly Gly Arg Arg Arg Gly Ala Gly Thr Pro Ala Ser Glu Asp  
 450 455 460  
 Gly Glu Arg Gly Pro Glu Pro Gly Ala Gln Gly Pro Glu Ser Trp Gly  
 465 470 475 480  
 15 Asp Ile Ala Thr Arg Ala Ala Ala Asp Val Xaa Xaa Xaa Xaa Xaa  
 485 490 495

## (2) INFORMATION FOR SEQ ID NO:132:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1186 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

30 Met Asp Thr Lys Pro Lys Thr Thr Thr Thr Val Lys Val Pro Pro Gly  
 1 5 10 15  
 Pro Met Gly Tyr Val Tyr Gly Arg Ala Cys Pro Ala Glu Gly Leu Glu  
 20 25 30  
 Leu Leu Ser Leu Leu Ser Ala Arg Ser Gly Asp Ala Asp Val Ala Val  
 35 35 40 45  
 Ala Pro Leu Ile Val Gly Leu Thr Val Glu Ser Gly Phe Glu Ala Asn  
 50 55 60  
 Val Ala Ala Val Val Gly Ser Arg Thr Thr Gly Leu Gly Gly Thr Ala  
 65 70 75 80  
 40 Val Ser Leu Lys Leu Met Pro Ser His Tyr Ser Pro Ser Val Tyr Val  
 85 90 95  
 Phe His Gly Gly Arg His Leu Ala Pro Ser Thr Gln Ala Pro Asn Leu  
 100 105 110  
 391

Thr Arg Leu Cys Glu Arg Ala Arg Arg His Phe Gly Phe Ser Asp Tyr  
 115 120 125  
 Ala Pro Arg Pro Cys Asp Leu Lys His Glu Thr Thr Gly Asp Ala Leu  
 130 135 140  
 5 Cys Glu Arg Leu Gly Leu Asp Pro Asp Arg Ala Leu Leu Tyr Leu Val  
 145 150 155 160  
 Ile Thr Glu Gly Phe Arg Glu Ala Val Cys Ile Ser Asn Thr Phe Leu  
 165 170 175  
 His Leu Gly Gly Met Asp Lys Val Thr Ile Gly Asp Ala Glu Val His  
 10 180 185 190  
 Arg Ile Pro Val Tyr Pro Leu Gln Met Phe Met Pro Asp Phe Ser Arg  
 195 200 205  
 Val Ile Ala Asp Pro Phe Asn Cys Asn His Arg Ser Ile Gly Glu Asn  
 210 215 220  
 15 Phe Asn Tyr Pro Leu Pro Phe Phe Asn Arg Pro Leu Ala Arg Leu Leu  
 225 230 235 240  
 Phe Glu Ala Val Val Gly Pro Ala Ala Val Arg Ala Arg Asn Val Asp  
 245 250 255  
 Ala Val Ala Arg Ala Ala Ala His Leu Ala Phe Asp Glu Asn His Glu  
 20 260 265 270  
 Gly Ala Ala Leu Pro Ala Asp Ile Thr Phe Thr Ala Phe Glu Ala Ser  
 275 280 285  
 Gln Gly Lys Pro Gln Arg Gly Ala Arg Asp Ala Gly Asn Lys Gly Pro  
 290 295 300  
 25 Ala Gly Gly Phe Glu Gln Arg Leu Ala Ser Val Met Ala Gly Asp Ala  
 305 310 315 320  
 Ala Leu Glu Ser Ile Val Ser Met Ala Val Phe Asp Glu Pro Pro Pro  
 325 330 335  
 Asp Ile Thr Thr Trp Pro Leu Leu Glu Gly Gln Glu Thr Pro Ala Ala  
 30 340 345 350  
 Arg Ala Gly Ala Val Gly Ala Tyr Leu Ala Arg Ala Ala Gly Leu Val  
 355 360 365  
 Gly Ala Met Val Phe Ser Thr Asn Ser Ala Leu His Leu Thr Glu Val  
 370 375 380  
 35 Asp Asp Ala Gly Pro Ala Asp Pro Lys Asp His Ser Lys Pro Ser Phe  
 385 390 395 400  
 Tyr Arg Phe Phe Leu Val Pro Gly Thr His Val Ala Ala Asn Pro Gln  
 405 410 415  
 Leu Asp Arg Glu Gly His Val Val Pro Gly Tyr Glu Gly Arg Pro Thr  
 40 420 425 430  
 Ala Pro Leu Val Gly Gly Thr Gln Glu Phe Ala Gly Glu His Leu Ala  
 435 440 445  
 Met Leu Cys Gly Phe Ser Pro Ala Leu Leu Ala Lys Met Leu Phe Tyr  
 392

	450		455		460	
	Leu Glu Arg Cys Asp Gly Gly Val Ile Val Gly Arg Gln Glu Met Asp					
	465		470		475	480
	Val Phe Arg Tyr Val Ala Asp Ser Gly Gln Thr Asp Val Pro Cys Asn					
5		485		490		495
	Leu Cys Thr Phe Glu Thr Arg His Ala Cys Ala His Thr Thr Leu Met					
		500		505		510
	Arg Leu Arg Ala Arg His Pro Lys Phe Ala Ser Ala Arg Ala Ile Gly					
		515		520		525
10	Val Phe Gly Thr Met Asn Ser Ala Tyr Ser Asp Cys Asp Val Leu Gly					
		530		535		540
	Asn Tyr Ala Ala Phe Ser Ala Leu Lys Arg Ala Asp Gly Ser Glu Asn					
		545		550		555
	Thr Arg Thr Ile Met Gln Glu Tyr Ala Ala Thr Glu Arg Val Met Ala					
15		565		570		575
	Glu Leu Glu Ala Leu Gln Tyr Val Asp Gln Ala Val Pro Thr Ala Leu					
		580		585		590
	Gly Arg Leu Glu Thr Ile Ile Gly Thr Arg Glu Ala Leu His Thr Val					
		595		600		605
20	Val Asn Asn Ile Lys Gln Leu Val Asp Arg Glu Val Glu Gln Leu Met					
		610		615		620
	Arg Asn Leu Ile Glu Gly Arg Asn Phe Lys Phe Arg Asp Gly Leu Ala					
		625		630		635
	Glu Ala Asn His Ala Met Ser Leu Ser Leu Asp Pro Tyr Thr Cys Gly					
25		645		650		655
	Pro Cys Pro Leu Leu Gln Leu Leu Ala Arg Arg Ser Asn Leu Ala Val					
		660		665		670
	Tyr Gln Asp Leu Ala Leu Ser Gln Cys His Gly Val Phe Ala Gly Gln					
		675		680		685
30	Ser Val Glu Gly Arg Asn Phe Arg Asn Gln Phe Gln Pro Val Leu Arg					
		690		695		700
	Arg Arg Val Met Asp Leu Phe Asn Asn Gly Phe Leu Ser Ala Lys Thr					
		705		710		715
	Leu Thr Val Ser Glu Gly Ala Ala Ile Cys Ala Pro Ser Leu Thr Ala					
35		725		730		735
	Gly Gln Thr Ala Pro Ala Glu Ser Ser Phe Glu Gly Asp Val Ala Arg					
		740		745		750
	Val Thr Leu Gly Phe Pro Lys Glu Leu Arg Val Lys Ser Arg Val Leu					
		755		760		765
40	Phe Ala Gly Ala Ser Ala Asn Ala Ser Glu Ala Ala Lys Ala Arg Val					
		770		775		780
	Ala Ser Leu Gln Ser Ala Tyr Gln Lys Pro Asp Lys Arg Val Asp Ile					
		785		790		795
						800



	Leu	Leu	Gly	Pro	Leu	Gly	Phe	Leu	Leu	Lys	Gln	Phe	His	Ala	Val	Ile	
						805				810						815	
	Phe	Pro	Asn	Gly	Lys	Pro	Pro	Gly	Ser	Asn	Gln	Pro	Asn	Pro	Gln	Trp	
				820					825					830			
5	Phe	Trp	Thr	Ala	Leu	Gln	Arg	Asn	Gln	Leu	Pro	Ala	Arg	Leu	Leu	Ser	
				835				840					845				
	Arg	Glu	Asp	Ile	Glu	Thr	Ile	Ala	Phe	Ile	Lys	Arg	Phe	Ser	Leu	Asp	
		850					855					860					
	Tyr	Gly	Ala	Ile	Asn	Phe	Ile	Asn	Leu	Ala	Pro	Asn	Asn	Val	Ser	Glu	
10	865					870					875					880	
	Leu	Ala	Met	Tyr	Tyr	Met	Ala	Asn	Gln	Ile	Leu	Arg	Tyr	Cys	Asp	His	
				885						890					895		
	Ser	Thr	Tyr	Phe	Ile	Asn	Thr	Leu	Thr	Ala	Val	Ile	Ala	Gly	Ser	Arg	
				900					905					910			
15	Arg	Pro	Pro	Gly	Val	Gln	Ala	Ala	Ala	Ala	Trp	Ala	Pro	Gln	Gly	Gly	
				915				920					925				
	Ala	Gly	Leu	Glu	Ala	Gly	Ala	Arg	Ala	Leu	Met	Asp	Ser	Leu	Asp	Ala	
		930				935					940						
	His	Pro	Gly	Ala	Trp	Thr	Ser	Met	Phe	Ala	Ser	Cys	Asn	Leu	Leu	Arg	
20	945					950					955					960	
	Pro	Val	Met	Ala	Ala	Arg	Pro	Met	Val	Val	Leu	Gly	Leu	Ser	Ile	Ser	
				965						970					975		
	Lys	Tyr	Tyr	Gly	Met	Ala	Gly	Asn	Asp	Arg	Val	Phe	Gln	Ala	Gly	Asn	
				980					985					990			
25	Trp	Ala	Ser	Leu	Leu	Gly	Gly	Lys	Asn	Ala	Cys	Pro	Leu	Leu	Ile	Phe	
		995						1000					1005				
	Asp	Arg	Thr	Arg	Lys	Phe	Val	Leu	Ala	Cys	Pro	Arg	Ala	Gly	Phe	Val	
		1010					1015					1020					
	Cys	Ala	Ala	Ser	Ser	Leu	Gly	Gly	Gly	Ala	His	Glu	His	Ser	Leu	Cys	
30	1025					1030					1035					104	
	Glu	Gln	Leu	Arg	Gly	Ile	Ile	Ala	Glu	Gly	Gly	Ala	Ala	Val	Ala	Ser	
				1045						1050					1055		
	Ser	Val	Phe	Val	Ala	Thr	Val	Lys	Ser	Leu	Gly	Pro	Arg	Thr	Gln	Gln	
				1060					1065					1070			
35	Leu	Gln	Ile	Glu	Asp	Trp	Leu	Ala	Leu	Leu	Glu	Asp	Glu	Tyr	Leu	Ser	
		1075						1080					1085				
	Glu	Glu	Met	Met	Glu	Phe	Thr	Thr	Arg	Ala	Leu	Glu	Arg	Gly	His	Gly	
		1090					1095					1100					
	Glu	Trp	Ser	Thr	Asp	Ala	Ala	Leu	Glu	Val	Ala	His	Glu	Ala	Glu	Ala	
40																	

1140 1145 1150  
 Ala Ala Gly Ala Ala Gly Val Ala Arg Lys Arg Ala Phe His Gly Asp  
 1155 1160 1165  
 Asp Pro Phe Gly Glu Gly Pro Pro Glu Lys Lys Asp Leu Thr Leu Asp  
 5 1170 1175 1180  
 Met Leu  
 1185

## (2) INFORMATION FOR SEQ ID NO:133:

10

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1228 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single.  
 (D) TOPOLOGY: linear

15

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

20

Met Phe Cys Ala Ala Gly Gly Pro Thr Ser Pro Gly Gly Lys Ser Ala  
 1 5 10 15  
 Ala Arg Ala Ala Ser Gly Phe Phe Ala Pro His Asn Pro Arg Gly Ala  
 20 25 30  
 25 Thr Gln Thr Ala Pro Pro Pro Cys Arg Arg Gln Asn Phe Tyr Asn Pro  
 35 40 45  
 His Leu Ala Gln Thr Gly Thr Gln Pro Lys Ala Pro Gly Pro Ala Gln  
 50 55 60  
 Arg His Thr Tyr Tyr Ser Glu Cys Asp Glu Phe Arg Phe Ile Ala Pro  
 30 65 70 75 80  
 Arg Ser Leu Asp Glu Asp Ala Pro Ala Glu Gln Arg Thr Gly Val His  
 85 90 95  
 Asp Gly Arg Leu Arg Arg Ala Pro Lys Val Tyr Cys Gly Gly Asp Glu  
 100 105 110  
 35 Arg Asp Val Leu Arg Val Gly Pro Glu Gly Phe Trp Pro Arg Arg Leu  
 115 120 125  
 Arg Leu Trp Gly Gly Ala Asp His Ala Pro Glu Gly Phe Asp Pro Thr  
 130 135 140  
 Val Thr Val Phe His Val Tyr Asp Ile His Val Glu His Ala Tyr Ser  
 40 145 150 155 160  
 Met Arg Ala Ala Gln Leu His Glu Arg Phe Met Asp Ala Ile Thr Pro  
 165 170 175  
 Ala Gly Thr Val Ile Thr Leu Leu Gly Leu Thr Pro Glu Gly His Arg

		180		185		190	
	Val	Ala	Val	His	Val	Tyr	Gly Thr Arg Gln Tyr Phe Tyr Met Asn Lys
		195		200		205	
	Ala	Glu	Val	Asp	Arg	His	Leu Gln Cys Arg Ala Pro Arg Asp Leu Cys
5		210		215		220	
	Glu	Arg	Leu	Ala	Ala	Ala	Leu Arg Glu Ser Pro Gly Ala Ser Phe Arg
	225			230		235	240
	Gly	Ile	Ser	Ala	Asp	His	Phe Glu Ala Glu Val Val Glu Arg Ala Asp
				245		250	255
10	Val	Tyr	Tyr	Tyr	Glu	Trp	Thr Leu Tyr Tyr Arg Val Phe Val Arg Ser
		260		265		270	
	Gly	Arg	Ala	Tyr	Leu	Cys	Asp Asn Phe Cys Pro Ala Ile Arg Lys Tyr
		275		280		285	
	Glu	Gly	Gly	Val	Asp	Ala	Thr Thr Arg Phe Ile Leu Asp Asn Pro Gly
15		290		295		300	
	Phe	Val	Thr	Phe	Gly	Trp	Tyr Arg Leu Lys Pro Gly Arg Gly Asn Ala
	305			310		315	320
	Pro	Ala	Gln	Pro	Arg	Pro	Pro Thr Ala Phe Gly Thr Ser Ser Asp Val
				325		330	335
20	Glu	Phe	Asn	Cys	Thr	Ala	Asp Asn Leu Ala Val Glu Gly Ala Met Cys
		340		345		350	
	Asp	Leu	Pro	Ala	Tyr	Lys	Leu Met Cys Phe Asp Ile Glu Cys Lys Ala
		355		360		365	
	Gly	Gly	Glu	Asp	Glu	Leu	Ala Phe Pro Val Ala Glu Arg Pro Glu Asp
25		370		375		380	
	Leu	Val	Ile	Gln	Ile	Ser	Cys Leu Leu Tyr Asp Leu Ser Thr Thr Ala
	385			390		395	400
	Leu	Glu	His	Ile	Leu	Leu	Phe Ser Leu Gly Ser Cys Asp Leu Pro Glu
				405		410	415
30	Ser	His	Leu	Ser	Asp	Leu	Ala Ser Arg Gly Leu Pro Ala Pro Val Val
		420		425		430	
	Leu	Glu	Phe	Asp	Ser	Glu	Phe Glu Met Leu Leu Ala Phe Met Thr Phe
		435		440		445	
	Val	Lys	Gln	Tyr	Gly	Pro	Glu Phe Val Thr Gly Tyr Asn Ile Ile Asn
35		450		455		460	
	Phe	Asp	Trp	Pro	Phe	Val	Leu Thr Lys Leu Thr Glu Ile Tyr Lys Val
	465			470		475	480
	Pro	Leu	Asp	Gly	Tyr	Gly	Arg Met Asn Gly Arg Gly Val Phe Arg Val
				485		490	495
40	Trp	Asp	Ile	Gly	Gln	Ser	His Phe Gln Lys Arg Ser Lys Ile Lys Val
		500		505		510	
	Asn	Gly	Met	Val	Asn	Ile	Asp Met Tyr Gly Ile Ile Thr Asp Lys Val
		515		520		525	

Lys Leu Ser Ser Tyr Lys Leu Asn Ala Val Ala Glu Ala Val Leu Lys  
 530 535 540  
 Asp Lys Lys Lys Asp Leu Ser Tyr Arg Asp Ile Pro Ala Tyr Tyr Ala  
 545 550 555 560  
 5 Ser Gly Pro Ala Gln Arg Gly Val Ile Gly Glu Tyr Cys Val Gln Asp  
 565 570 575  
 Ser Leu Leu Val Gly Gln Leu Phe Phe Lys Phe Leu Pro His Leu Glu  
 580 585 590  
 Leu Ser Ala Val Ala Arg Leu Ala Gly Ile Asn Ile Thr Arg Thr Ile  
 10 595 600 605  
 Tyr Asp Gly Gln Gln Ile Arg Val Phe Thr Cys Leu Leu Arg Leu Ala  
 610 615 620  
 Gly Gln Lys Gly Phe Ile Leu Pro Asp Thr Gln Gly Arg Phe Arg Gly  
 625 630 635 640  
 15 Leu Asp Lys Glu Ala Pro Lys Arg Pro Ala Val Pro Arg Gly Glu Gly  
 645 650 655  
 Glu Arg Pro Gly Asp Gly Asn Gly Asp Glu Asp Lys Asp Asp Asp Glu  
 660 665 670  
 Asp Gly Asp Glu Asp Gly Asp Glu Arg Glu Glu Val Ala Arg Glu Thr  
 20 675 680 685  
 Gly Gly Arg His Val Gly Tyr Gln Gly Ala Arg Val Leu Asp Pro Thr  
 690 695 700  
 Ser Gly Phe His Val Asp Pro Val Val Val Phe Asp Phe Ala Ser Leu  
 705 710 715 720  
 25 Tyr Pro Ser Ile Ile Gln Ala His Asn Leu Cys Phe Ser Thr Leu Ser  
 725 730 735  
 Leu Arg Pro Glu Ala Val Ala His Leu Glu Ala Asp Arg Asp Tyr Leu  
 740 745 750  
 Glu Ile Glu Val Gly Gly Arg Arg Leu Phe Phe Val Lys Ala His Val  
 30 755 760 765  
 Arg Glu Ser Leu Leu Ser Ile Leu Leu Arg Asp Trp Leu Ala Met Arg  
 770 775 780  
 Lys Gln Ile Arg Ser Arg Ile Pro Gln Ser Thr Pro Glu Glu Ala Val  
 785 790 795 800  
 35 Leu Leu Asp Lys Gln Gln Ala Ala Ile Lys Val Val Cys Asn Ser Val  
 805 810 815  
 Tyr Gly Phe Thr Gly Val Gln His Gly Leu Leu Pro Cys Leu His Val  
 820 825 830  
 Ala Ala Thr Val Thr Thr Ile Gly Arg Glu Met Leu Leu Ala Thr Arg  
 40 835 840 845  
 Ala Tyr Val His Ala Arg Trp Ala Glu Phe Asp Gln Leu Leu Ala Asp  
 850 855 860  
 Phe Pro Glu Ala Ala Gly Met Arg Ala Pro Gly Pro Tyr Ser Met Arg  
 397

	865		870		875		880									
	Ile	Ile	Tyr	Gly	Asp	Thr	Asp	Ser	Ile	Phe	Val	Leu	Cys	Arg	Gly	Leu
				885					890						895	
	Thr	Ala	Ala	Gly	Leu	Val	Ala	Met	Gly	Asp	Lys	Met	Ala	Ser	His	Arg
5			900					905						910		
	Ala	Leu	Phe	Leu	Pro	Pro	Ile	Lys	Leu	Glu	Cys	Glu	Lys	Thr	Phe	Thr
			915					920						925		
	Lys	Leu	Leu	Leu	Ile	Ala	Lys	Lys	Lys	Tyr	Ile	Gly	Val	Ile	Cys	Gly
		930					935					940				
10	Gly	Lys	Met	Leu	Ile	Lys	Gly	Val	Asp	Leu	Val	Arg	Lys	Asn	Asn	Cys
		945				950					955				960	
	Ala	Phe	Ile	Asn	Arg	Thr	Ser	Arg	Ala	Leu	Val	Asp	Leu	Leu	Phe	Tyr
				965					970						975	
	Asp	Asp	Thr	Val	Ser	Gly	Ala	Ala	Ala	Ala	Glu	Arg	Pro	Ala	Glu	Glu
15			980						985					990		
	Trp	Leu	Ala	Arg	Pro	Leu	Pro	Glu	Gly	Leu	Gln	Ala	Phe	Gly	Ala	Val
			995					1000						1005		
	Leu	Val	Asp	Ala	His	Arg	Arg	Ile	Thr	Asp	Pro	Glu	Arg	Asp	Ile	Gln
		1010				1015					1020					
20	Asp	Phe	Val	Leu	Thr	Ala	Glu	Leu	Ser	Arg	His	Pro	Arg	Ala	Tyr	Thr
		1025				1030					1035				104	
	Asn	Lys	Arg	Leu	Ala	His	Leu	Thr	Val	Tyr	Tyr	Lys	Leu	Met	Ala	Arg
				1045						1050					1055	
	Arg	Ala	Gln	Val	Pro	Ser	Ile	Lys	Asp	Arg	Ile	Pro	Tyr	Val	Ile	Val
25			1060						1065					1070		
	Ala	Gln	Thr	Arg	Glu	Val	Glu	Glu	Thr	Val	Ala	Arg	Leu	Ala	Ala	Leu
			1075					1080					1085			
	Arg	Glu	Leu	Asp	Ala	Ala	Ala	Pro	Gly	Asp	Glu	Pro	Ala	Pro	Pro	Ala
		1090				1095					1100					
30	Ala	Leu	Pro	Ser	Pro	Ala	Lys	Arg	Pro	Arg	Glu	Thr	Pro	Ser	His	Ala
		1105				1110					1115				112	
	Asp	Pro	Pro	Gly	Gly	Ala	Ser	Lys	Pro	Arg	Lys	Leu	Leu	Val	Ser	Glu
				1125					1130					1135		
	Leu	Ala	Glu	Asp	Pro	Gly	Tyr	Ala	Ile	Arg	Val	Pro	Leu	Asn	Thr	Asp
35			1140						1145					1150		
	Tyr	Tyr	Phe	Ser	His	Leu	Leu	Gly	Ala	Ala	Cys	Val	Thr	Phe	Lys	Ala
			1155						1160					1165		
	Leu	Phe	Gly	Asn	Asn	Ala	Lys	Ile	Thr	Glu	Ser	Leu	Leu	Lys	Arg	Phe
		1170				1175					1180					
40	Ile	Pro	Glu	Thr	Trp	His	Pro	Pro	Asp	Asp	Val	Ala	Ala	Arg	Leu	Arg
		1185				1190					1195				120	
	Ala	Ala	Gly	Phe	Gly	Pro	Ala	Gly	Ala	Gly	Ala	Thr	Ala	Glu	Glu	Thr
				1205					1210					1215		

Arg Arg Met Leu His Arg Ala Phe Asp Thr Leu Ala  
 1220 1225

## (2) INFORMATION FOR SEQ ID NO:134:

5

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 303 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

10

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

15

Met Tyr Asp Ile Ala Pro Arg Arg Ser Gly Ser Arg Pro Gly Pro Gly  
 1 5 10 15  
 Arg Asp Lys Thr Arg Arg Arg Ser Arg Phe Ser Ala Ala Gly Asn Pro  
 20 25 30  
 Gly Val Glu Arg Arg Ala Ser Arg Lys Ser Leu Pro Ser His Ala Arg  
 35 40 45  
 Arg Leu Glu Leu Cys Leu His Glu Arg Arg Arg Tyr Arg Gly Phe Phe  
 50 55 60  
 Ala Ala Gln Thr Pro Ser Glu Glu Ile Ala Ile Val Arg Ser Leu Ser  
 25 65 70 75 80  
 Val Pro Leu Val Lys Thr Thr Pro Val Ser Leu Pro Phe Ser Leu Asp  
 85 90 95  
 Gln Thr Val Ala Asp Asn Cys Leu Thr Leu Ser Gly Met Gly Tyr Tyr  
 100 105 110  
 30 Leu Gly Ile Gly Gly Cys Cys Pro Ala Cys Ser Ala Gly Asp Gly Arg  
 115 120 125  
 Leu Ala Thr Val Ser Arg Glu Ala Leu Ile Leu Ala Phe Val Gln Gln  
 130 135 140  
 Ile Asn Thr Ile Phe Glu His Arg Thr Phe Leu Ala Ser Leu Val Val  
 35 145 150 155 160  
 Leu Ala Asp Arg His Ser Thr Pro Leu Gln Asp Leu Leu Ala Asp Thr  
 165 170 175  
 Leu Gly Gln Pro Glu Leu Phe Phe Val His Thr Ile Leu Arg Gly Gly  
 180 185 190  
 40 Gly Ala Cys Asp Pro Arg Phe Leu Phe Tyr Pro Asp Pro Thr Tyr Gly  
 195 200 205  
 Gly His Met Leu Tyr Val Ile Phe Pro Gly Thr Ser Ala His Leu His  
 210 215 220

Tyr Arg Leu Ile Asp Arg Met Leu Thr Ala Cys Pro Gly Tyr Arg Phe  
 225 230 235 240  
 Ala Ala His Val Trp Gln Ser Thr Phe Val Leu Val Val Arg Arg Asn  
 245 250 255  
 5 Ala Glu Lys Pro Ala Asp Ala Glu Ile Pro Thr Val Ser Ala Ala Asp  
 260 265 270  
 Ile Tyr Cys Lys Met Arg Asp Ile Ser Phe Asp Gly Gly Leu Met Leu  
 275 280 285  
 Glu Tyr Gln Arg Leu Tyr Ala Thr Phe Asp Glu Phe Pro Pro Pro  
 10 290 295 300

## (2) INFORMATION FOR SEQ ID NO:135:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 597 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

Val Arg Pro Ala Arg Pro Ala Met Ala Thr Ser Ala Pro Gly Val Pro  
 25 1 5 10 15  
 Ser Ser Ala Ala Val Arg Glu Glu Ser Pro Gly Ser Ser Trp Lys Glu  
 20 25 30  
 Gly Ala Phe Glu Arg Pro Tyr Val Ala Phe Asp Pro Asp Leu Leu Ala  
 35 40 45  
 30 Leu Asn Glu Ala Leu Cys Ala Glu Leu Leu Ala Ala Cys His Val Val  
 50 55 60  
 Gly Val Pro Pro Ala Ser Ala Leu Asp Glu Asp Val Glu Ser Asp Val  
 65 70 75 80  
 Ala Pro Ala Pro Pro Arg Pro Arg Gly Ala Ala Arg Glu Ala Ser Gly  
 35 85 90 95  
 Gly Arg Gly Pro Gly Ser Arg Pro Pro Ala Asp Pro Thr Ala Glu Gly  
 100 105 110  
 Leu Leu Asp Thr Gly Pro Phe Ala Ala Ala Ser Val Asp Thr Phe Ala  
 115 120 125  
 40 Leu Asp Arg Pro Cys Leu Val Cys Arg Thr Ile Glu Leu Tyr Lys Gln  
 130 135 140  
 Ala Tyr Arg Leu Ser Pro Gln Trp Val Ala Asp Tyr Ala Phe Leu Cys  
 145 150 155 160

Ala Lys Cys Leu Gly Ala Pro His Cys Ala Ala Ser Ile Phe Val Ala  
 165 170 175  
 Ala Phe Glu Phe Val Tyr Val Met Asp His His Phe Leu Arg Thr Lys  
 180 185 190  
 5 Lys Ala Thr Leu Val Gly Ser Phe Ala Arg Phe Ala Leu Thr Ile Asn  
 195 200 205  
 Asp Ile His Arg His Phe Phe Leu His Cys Cys Phe Arg Thr Asp Gly  
 210 215 220  
 Gly Val Pro Gly Arg His Ala Gln Lys Gln Pro Arg Pro Thr Pro Ser  
 10 225 230 235 240  
 Pro Gly Ala Ala Lys Val Gln Tyr Ser Asn Tyr Ser Phe Leu Ala Gln  
 245 250 255  
 Ser Ala Thr Arg Ala Leu Ile Gly Thr Leu Ala Ser Gly Gly Asp Asp  
 260 265 270  
 15 Gly Ala Gly Ala Gly Gly Gly Ser Gly Thr Gln Pro Ser Leu Thr Thr  
 275 280 285  
 Ala Leu Met Asn Trp Lys Asp Cys Ala Arg Leu Leu Asp Cys Thr Glu  
 290 295 300  
 Gly Lys Arg Gly Gly Gly Asp Ser Cys Cys Thr Arg Ala Ala Ala Arg  
 20 305 310 315 320  
 Asn Gly Glu Phe Glu Ala Ala Ala Gly Ala Gln Gly Gly Glu Pro Glu  
 325 330 335  
 Thr Trp Ala Tyr Ala Asp Leu Ile Leu Leu Leu Ala Gly Thr Pro  
 340 345 350  
 25 Ala Val Trp Glu Ser Gly Pro Arg Leu Arg Ala Ala Ala Asp Ala Arg  
 355 360 365  
 Arg Ala Ala Val Ser Glu Ser Trp Glu Ala His Arg Gly Ala Arg Met  
 370 375 380  
 Arg Asp Ala Ala Pro Arg Phe Ala Gln Phe Ala Glu Pro Lys Ala Gln  
 30 385 390 395 400  
 Pro Asp Leu Asp Leu Gly Pro Leu Met Ala Thr Val Leu Lys His Gly  
 405 410 415  
 Arg Gly Arg Gly Arg Thr Gly Gly Glu Cys Leu Leu Cys Asn Leu Leu  
 420 425 430  
 35 Leu Val Arg Ala Tyr Trp Leu Ala Met Arg Arg Leu Arg Ala Ser Val  
 435 440 445  
 Val Arg Tyr Ser Glu Asn Asn Thr Ser Leu Phe Asp Cys Ile Val Pro  
 450 455 460  
 Val Val Asp Gln Leu Glu Ala Asp Pro Glu Ala Gln Pro Gly Asp Gly  
 40 465 470 475 480  
 Gly Arg Phe Val Ser Leu Leu Arg Ala Ala Gly Pro Glu Ala Ile Phe  
 485 490 495  
 Lys His Met Phe Cys Asp Pro Met Cys Ala Ile Thr Glu Met Glu Val  
 401



```

                    500                    505                    510
Asp Pro Trp Val Leu Phe Gly His Pro Arg Ala Asp His Arg Asp Glu
                    515                    520                    525
Leu Gln Leu His Lys Ala Lys Leu Ala Cys Gly Asn Glu Phe Glu Gly
5      530                    535                    540
Arg Val Cys Ile Ala Leu Arg Ala Leu Ile Tyr Thr Phe Lys Thr Tyr
545                    550                    555                    560
Gln Val Phe Val Pro Lys Pro Thr Ala Thr Phe Val Arg Glu Ala Gly
                    565                    570                    575
10   Ala Leu Leu Arg Arg His Ser Ile Ser Leu Leu Ser Leu Glu His Thr
                    580                    585                    590
Leu Cys Thr Tyr Val
                    595

```

15 (2) INFORMATION FOR SEQ ID NO:136:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 128 amino acids  
 (B) TYPE: amino acid  
 20 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

```

Met Ala Gly Arg Ala Gly Arg Trp Arg Thr Leu Arg Asp Ala Ile Pro
  1              5              10              15
Asp Cys Ala Leu Arg Ser Gln Thr Leu Glu Ser Leu Asp Ala Arg Tyr
30              20              25              30
Val Ser Arg Asp Gly Ala Gly Asp Ala Ala Val Trp Phe Glu Asp Met
              35              40              45
Thr Pro Ala Glu Leu Glu Val Ile Phe Pro Thr Thr Asp Ala Lys Leu
              50              55              60
35   Asn Tyr Leu Ser Arg Thr Gln Arg Leu Ala Ser Leu Leu Thr Tyr Ala
              65              70              75              80
Gly Pro Ile Lys Ala Pro Asp Gly Pro Ala Ala Pro His Thr Gln Asp
              85              90              95
40   Thr Ala Cys Val His Gly Glu Leu Asp Ala Thr Glu Arg Glu Arg Phe
              100             105             110
Ala Ala Val Ile Asn Arg Phe Leu Asp Leu His Gln Ile Leu Arg Gly
              115             120             125

```

## (2) INFORMATION FOR SEQ ID NO:137:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 274 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Met Ala Gly Met Gly Lys Pro Tyr Gly Gly Arg Pro Gly Asp Ala Phe  
 1 5 10 15  
 15 Glu Gly Leu Val Gln Arg Ile Arg Leu Ile Val Pro Thr Thr Leu Arg  
 20 25 30  
 Gly Gly Gly Gly Glu Ser Gly Pro Tyr Ser Pro Ser Asn Pro Pro Ser  
 35 40 45  
 Arg Cys Ala Phe Gln Phe His Gly Gln Asp Gly Ser Asp Glu Ala Phe  
 20 50 55 60  
 Pro Ile Glu Tyr Val Leu Arg Leu Met Asn Asp Trp Ala Asp Val Pro  
 65 70 75 80  
 Cys Asn Pro Tyr Leu Arg Val Gln Asn Thr Gly Val Ser Val Leu Phe  
 85 90 95  
 25 Gln Gly Phe Phe Asn Arg Pro His Gly Ala Pro Gly Gly Ala Ile Thr  
 100 105 110  
 Ala Glu Gln Thr Asn Val Ile Leu His Ser Thr Glu Thr Thr Gly Leu  
 115 120 125  
 Ser Leu Gly Asp Leu Asp Asp Val Lys Gly Arg Leu Gly Leu Asp Ala  
 30 130 135 140  
 Arg Pro Met Met Ala Ser Met Trp Ile Ser Cys Phe Val Arg Met Pro  
 145 150 155 160  
 Arg Val Gln Leu Ala Phe Arg Phe Met Gly Pro Glu Asp Ala Val Arg  
 165 170 175  
 35 Thr Arg Arg Ile Leu Cys Arg Ala Ala Glu Gln Ala Arg Arg Arg Arg  
 180 185 190  
 Ser Arg Arg Ser Gln Asp Asp Tyr Gly Ala Val Ala Val Ala Ala Ala  
 195 200 205  
 His His Ser Ser Gly Ala Pro Gly Pro Gly Val Ala Ala Ser Gly Pro  
 40 210 215 220  
 Pro Ala Pro Pro Gly Arg Gly Pro Ala Arg Pro Trp His Gln Ala Val  
 225 230 235 240  
 Gln Leu Phe Arg Ala Pro Arg Pro Gly Pro Pro Ala Leu Leu Leu Leu

Val Ala Gly Leu Phe Leu Gly Ala Ala Ile Trp Trp Ala Val Gly Ala  
Arg Leu

5

(2) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 112 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

20	Met	Ala	Ala	Pro	Gln	Phe	His	Arg	Pro	Ser	Thr	Ile	Thr	Ala	Asp	Asn
	1				5					10					15	
	Val	Arg	Ala	Leu	Gly	Met	Arg	Gly	Leu	Val	Leu	Ala	Thr	Asn	Asn	Ala
				20					25					30		
25	Gln	Phe	Ile	Met	Asp	Asn	Ser	Tyr	Pro	His	Pro	His	Gly	Thr	Gln	Gly
			35					40					45			
	Ala	Val	Arg	Glu	Phe	Leu	Arg	Gly	Gln	Ala	Ala	Ala	Leu	Thr	Asp	Leu
			50				55					60				
30	Gly	Val	Thr	His	Ala	Asn	Asn	Thr	Phe	Ala	Pro	Gln	Pro	Met	Phe	Ala
	65					70					75				80	
	Gly	Asp	Ala	Ala	Ala	Glu	Trp	Leu	Arg	Pro	Ser	Phe	Gly	Leu	Lys	Arg
					85					90					95	
30	Thr	Tyr	Ser	Pro	Phe	Val	Val	Arg	Asp	Pro	Lys	Thr	Pro	Ser	Thr	Pro
				100					105					110		

(2) INFORMATION FOR SEQ ID NO:139:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 837 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

```

      CCCGCTAGTC TGGGGGCGAG GTGCTGCAGG ACCGAGTAGA GGATGGAAAA AACGTCTCGG      60
      TCGTAAACCA CGACCGAGCG GGGTCCGATG CAGCCGTCGG GGCCGCTCTC GACGATGGCC      120
      ACCAGCGGAC AGTCGGAGTT GTACGTGAGG TACACGCCCC GCGGGTAGCG GTACAGACCT      180
5      TCGGAGGTCG GGC GGCTGCA GTCGGGGCGG CGCAACTCAA GCTCCCCGCA CCGGTAGACC      240
      GACGCAAAGA GTGTGGTGGC GATAATGAGC TCGCGAATAT ATCGCCAGGC GGCGCGCTGG      300
      GTGGGCGTGA TTCCGGAAAC ACCGTCAAAA CAGTAGAACT TTTGAAACTC GCTGACGGCC      360
      CAATCAGCGC CCGAACCCCC CGCGCCCATG ATGAAGCGGG CGAGTTCCTC CTTGAGGTGC      420
      GGCAGGAGCC CCACGTCTCT GACGCTGTAG TACAGCGCGG TGTTGGGGGG CTGGGCGAAG      480
10     CTGTGGGTGG AGTGGTCGAA CAGGGGCCCC TTGACGAGCT CGAAGAAGCG ATGGGTGATG      540
      CTGGGGAGCA GGGCCGGGTC CACCTGGTGG CGCAGCAGCG ACGCTCGCAT GAACCGGTGC      600
      GCGTCAAACA CGCCCGGGGC GGC GCGGTTG TCGATGACCG TGCCCGCGCC CGCCGTCAGG      660
      GCGCAGAAGC GCGCGCGCGC CGCGAAGCCG TTGGCGACCG CGGCGAAGGT CGCGGGCAGC      720
      ACCTCGCCGT GGACGCTGAC CCGCAGCATC TTCTCGAGCT CCCC GCGCTG CTCGCGCACG      780
15     CAGCGCCCGA GGCTGGCCAG CGACCGCTTG GTCAGGCGGT CCGCGTACAG CCGCCG      837

```

## (2) INFORMATION FOR SEQ ID NO:140:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 278 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 25 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

```

      Arg Arg Leu Tyr Ala Asp Arg Leu Thr Lys Arg Ser Leu Ala Ser Leu
30      1           5           10          15
      Gly Arg Cys Val Arg Glu Gln Arg Gly Glu Leu Glu Lys Met Leu Arg
           20           25           30
      Val Ser Val His Gly Glu Val Leu Pro Ala Thr Phe Ala Ala Val Ala
           35           40           45
35     Asn Gly Phe Ala Ala Arg Ala Arg Phe Cys Ala Leu Thr Ala Gly Ala
           50           55           60
      Gly Thr Val Ile Asp Asn Arg Ala Ala Pro Gly Val Phe Asp Ala His
           65           70           75           80
      Arg Phe Met Arg Ala Ser Leu Leu Arg His Gln Val Asp Pro Ala Leu
40           85           90           95
      Leu Pro Ser Ile Thr Phe Phe Glu Leu Val Asn Gly Pro Leu Phe Asp
           100          105          110
      His Ser Thr His Ser Phe Ala Gln Pro Pro Asn Thr Ala Leu Tyr Tyr

```

115 120 125  
 Ser Val Glu Asn Val Gly Leu Leu Pro His Leu Lys Glu Glu Leu Ala  
 130 135 140  
 Arg Phe Ile Met Gly Ala Gly Gly Ser Gly Ala Asp Trp Ala Val Ser  
 5 145 150 155 160  
 Glu Phe Gln Lys Phe Tyr Cys Phe Asp Gly Val Ser Gly Ile Thr Pro  
 165 170 175  
 Thr Gln Arg Ala Ala Trp Arg Tyr Ile Arg Glu Leu Ile Ile Ala Thr  
 180 185 190  
 10 Thr Leu Phe Ala Ser Val Tyr Arg Cys Gly Glu Leu Glu Leu Arg Arg  
 195 200 205  
 Pro Asp Cys Ser Arg Pro Thr Ser Glu Gly Arg Tyr Pro Pro Gly Val  
 210 215 220  
 Tyr Leu Thr Tyr Asn Ser Asp Cys Pro Leu Val Ala Ile Val Glu Ser  
 15 225 230 235 240  
 Gly Pro Asp Gly Cys Ile Gly Pro Arg Ser Val Val Val Tyr Asp Arg  
 245 250 255  
 Asp Val Phe Ser Ile Lys Val Leu Gln His Leu Ala Pro Arg Leu Ala  
 260 265 270  
 20 Gly Xaa Xaa Xaa Xaa Xaa  
 275

## (2) INFORMATION FOR SEQ ID NO:141:

- 25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2646 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

AAACAATACC AGAAGTCATG TGTATTTTGT AACATCGGTG TCTTTTATT TATACACAAG 60  
 35 CCCAGCTCCC CTCCCCTCCC TTAGAGCTCG TCTTCGTCTC CGGCCTCGTC CTCGTTGTGG 120  
 AGCGGAGAGT ACCTGGCTTT GTTGCGCTTG CGCAGAACCA TGTTGGTGAC CTTGGAGCTG 180  
 AGCAGGGCGC TCGTGCCCTT CTTTCTGGCC TTGTGTTCCG TGCGCTCCAT GGCCGACACC 240  
 AAAGCCATAT ATCGGATCAT TTCTCGGGCC TCGGCCAACT TGGCCTCGTC AAACCCGCCC 300  
 CCCTCCGCGC CTTCCCTCCC CTCCCAGCCC ACGCCCCCGG GGTCGGAAGT CTTGAGTTCC 360  
 40 TTGGTGGTGA GCGGATACAG GGCCTTCATG GGATTGCGTT GCAGTTGCAG GACGTAGCGG 420  
 AAGGCGAAGA AGGCCGCGAC CAGGCCGGCC AGGACCAGCA GCCCCACGGC AAGCGCCCCG 480  
 AAGGGGTTGG ACATAAAGGA GGACACGCCC GAGACGGCCG ACACCACGCC CCCCCTACT 540  
 CCCATGACTA CCTTGCCGAC CGCGCGCCCC AAGTCCCCCA TCCCCTCGAA GAACGCGCAC 600

	AGCCCCGCGA	ACATGGCGGC	GTTGGCGTCG	GCGCGGATGA	CCGTGTCGAT	GTCGGCAAAG	660
	CGCAGGTCGT	GCAGCTGGTT	GCGGCGCTGG	ACCTCCGTGT	AGTCCAGCAG	GCCGCTGTCC	720
	TTGATCTCGT	GGCGCGTGTA	GACCTCCAGG	GGCACAAACT	CGTGGTCCTC	CAGCATGGTG	780
	ATGTTCAAGT	CGATGAAGGT	GCTGACGGTG	GTGACGTCGG	CGCGACTCAG	CTGGTGAGAG	840
5	TACGCGTACT	CCTCGAAGTA	CACGTAGCCC	CCGCCGAAGA	TGAAGTAGCG	CCGGTGGGCC	900
	ACGGTGACAG	GCTCGAGCGC	GTCGCGGGTG	AGGCGCAGCT	CGTTGTTCTC	GCCCAGCTGC	960
	CCCTCGATCA	GCGGGCCCTG	GTCTTCGTAC	CGAAAGCTGA	CCAGGGGGCG	GCTGTAGCAC	1020
	GTCCCCGGCC	GCGAGCTGAC	GCGCATCGAG	TTCTGCACGA	TCACGTGTGC	CGGGGCGACG	1080
	GGCACGCACG	TGGAGACGGC	CATGACGTCT	CCGAGCATGC	GCGCGCTCAC	CCGCCGGCCG	1140
10	ACGGTGCGCG	AGGCGATGGC	GTTGGGGTTG	AGCTTGCGGG	CCTCGTTCCA	GAGAGTCAGC	1200
	TCGTGGTTCT	GCAGCTCGCA	CCACGCGACG	GCGATGCGCC	CCAGCATGTC	GTTACAGTGG	1260
	CGCTGTATGT	GGTTATACGT	AAACTGCAGC	CGGGCGAACT	CGATCGAGGA	GGTGGTCTTG	1320
	ATGCGCTCCA	CGGACGCGTT	GCGCTGGGC	GCCTCCCGCA	GTGGCGCGGG	CGTGGCATTC	1380
	CGGGGCTTGC	GGTCCTGCTC	CCGCATGTAC	TCCCGCACGT	ACAGCTCGGC	GAGCGTGTTC	1440
15	CTGAGGAGGG	GCTGGTACGC	GATGAGGAAG	CCCCCGTGG	CCAGGTAGTA	CTGCGGCTGG	1500
	CCCACCTTGA	TGTGCGTGGC	GTTGTACTTG	GCGCAAACA	TGCGGTCGAT	GGCCTCGCGG	1560
	GCATCCCGGC	CAATGCAGTC	GCCCAGGTGC	ACGCGCGAGA	GCGAGTACTG	GGTCAGGTTG	1620
	GTGGTGAAGG	TGGTCGAGAT	GGCGTCGGAG	GAGAAGCGGA	AGGAGCCGCC	GTACTCGGCG	1680
	CGGAGCATCT	CGTCCACCTC	CTGCCACTTG	GTGATGGTGC	AGACCGCCGG	TCGCTTCGGC	1740
20	ACCCAGTCCC	AGGCCACGGT	AAACTTGGGG	GTGCTCAGCA	AGTTGCGGGT	CGTCGGCGAC	1800
	GTGGCCCGGG	CCTTCGTGGT	GAGGTCGCGC	GCGTAGAAGC	CGTCGACCTG	CTTGAAGCGG	1860
	TCGCGCGCGT	AGCTGGTGTG	CTCGGTGTGC	GACCCCTCCC	GGTAGCCGTA	AAACGGGGAC	1920
	ATGTACACAA	AGTCGCCCCT	CGCCAACACA	AACTCATCGT	ACGGGTACAC	CGACCGCGCG	1980
	TCCACCTCCT	CGACGATGCA	GTTGACCGTC	GTGCCGTACC	GATGGAACGC	CTCCACCCGC	2040
25	GAGGGGTTGT	ACTTGAGGTC	GGTGGTGTGC	CACCCCGGCG	TCGTGCGCGT	GGCGACCTTC	2100
	GCCGCGCTTGA	GCTCCATGTC	GGTCTCGTGG	TCGTCCCGGT	GAAACGCGGT	GGTCTCCATG	2160
	TTGTTCGCA	CGTACTTGGC	CGTGAGCGG	CAGACCCCT	TGGCGTTAAT	CTTGTCGATC	2220
	ACCTCCTCGA	AGGGAACGGG	GGCGCGGTCC	TCGAATATCC	CCATAAACTG	GGAGTAGCGG	2280
	TGGCCGAACC	ACACCTGCGA	CACGGTCACG	TCTTTGTAGT	ACATGGTGGC	CTTGAATTTC	2340
30	TACGGGGCGA	TGTTCTCCTT	GAAGACCACC	GCGATGCCCT	CCGTGTAGTT	CTGCCCTTCC	2400
	GGGCGCGTCG	GGCAGCGGCG	CGGCTGCTCA	AACTGCACCA	CCGTGGCGCC	CGTCGGGGGC	2460
	GGGCACACGT	AAAACTGGGC	ATCGGCGTTC	TCGACCTTGA	TTTCCCGCAG	GTGCGCGCGC	2520
	AGCGTGGCGT	GGCCGGCGGC	GACGGTCGCG	TTGGCGTCGG	GGGGCGGGGT	CGCCTCGGGC	2580
	CGCTTGGGCG	GCTTTTTTGGT	TTTCCGCTTC	CGGGCCTTGG	TGGTCGCGGG	GCTCGGGACG	2640
35	GGGGG						2646

## (2) INFORMATION FOR SEQ ID NO:142:

## (i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 846 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

5  
 Pro Pro Val Pro Ser Pro Ala Thr Thr Lys Ala Arg Lys Arg Lys Thr  
 1 5 10 15  
 Lys Lys Pro Pro Lys Arg Pro Glu Ala Thr Pro Pro Pro Asp Ala Asn  
 20 25 30  
 10 Ala Thr Val Ala Ala Gly His Ala Thr Leu Arg Ala His Leu Arg Glu  
 35 40 45  
 Ile Lys Val Glu Asn Ala Asp Ala Gln Phe Tyr Val Cys Pro Pro Pro  
 50 55 60  
 Thr Gly Ala Thr Val Val Gln Phe Glu Gln Pro Arg Arg Cys Pro Trp  
 15 65 70 75 80  
 Glu Gly Gln Asn Tyr Thr Glu Gly Ile Ala Val Val Phe Lys Glu Asn  
 85 90 95  
 Ile Ala Pro Tyr Lys Phe Lys Ala Thr Met Tyr Tyr Lys Asp Val Thr  
 100 105 110  
 20 Val Ser Gln Val Trp Phe Gly His Arg Tyr Ser Gln Phe Met Gly Ile  
 115 120 125  
 Phe Glu Asp Arg Ala Pro Val Pro Phe Glu Glu Val Ile Asp Lys Ile  
 130 135 140  
 Asn Ala Lys Gly Val Cys Arg Ser Thr Ala Lys Tyr Val Arg Asn Asn  
 25 145 150 155 160  
 Met Thr Ala Phe His Arg Asp Asp His Glu Thr Asp Met Glu Leu Lys  
 165 170 175  
 Pro Ala Lys Val Ala Thr Arg Thr Ser Arg Gly Trp His Thr Thr Asp  
 180 185 190  
 30 Leu Lys Tyr Asn Pro Ser Arg Val Glu Ala Phe His Arg Tyr Gly Thr  
 195 200 205  
 Thr Val Asn Cys Ile Val Glu Glu Val Asp Ala Arg Ser Val Tyr Pro  
 210 215 220  
 Tyr Asp Glu Phe Val Leu Ala Thr Gly Asp Phe Val Tyr Met Ser Pro  
 35 225 230 235 240  
 Phe Tyr Gly Tyr Arg Glu Gly Ser His Thr Glu His Thr Ser Tyr Ala  
 245 250 255  
 Ala Asp Arg Phe Lys Gln Val Asp Gly Phe Tyr Ala Arg Asp Leu Thr  
 260 265 270  
 40 Thr Lys Ala Arg Ala Thr Ser Pro Thr Thr Arg Asn Leu Leu Thr Thr  
 275 280 285  
 Pro Lys Phe Thr Val Ala Trp Asp Trp Val Pro Lys Arg Pro Ala Val  
 290 295 300

Cys Thr Met Thr Lys Trp Gln Glu Val Asp Glu Met Leu Arg Ala Glu  
 305 310 315 320  
 Tyr Gly Gly Ser Phe Arg Phe Ser Ser Asp Ala Ile Ser Thr Thr Phe  
 325 330 335  
 5 Thr Thr Asn Leu Thr Gln Tyr Ser Leu Ser Arg Val Asp Leu Gly Asp  
 340 345 350  
 Cys Ile Gly Arg Asp Ala Arg Glu Ala Ile Asp Arg Met Phe Ala Arg  
 355 360 365  
 10 Lys Tyr Asn Ala Thr His Ile Lys Val Gly Gln Pro Gln Tyr Tyr Leu  
 370 375 380  
 Ala Thr Gly Gly Phe Leu Ile Ala Tyr Gln Pro Leu Leu Ser Asn Thr  
 385 390 395 400  
 Leu Ala Glu Leu Tyr Val Arg Glu Tyr Met Arg Glu Gln Asp Arg Lys  
 405 410 415  
 15 Pro Arg Asn Ala Thr Pro Ala Pro Leu Arg Glu Ala Pro Ser Ala Asn  
 420 425 430  
 Ala Ser Val Glu Arg Ile Lys Thr Thr Ser Ser Ile Glu Phe Ala Arg  
 435 440 445  
 Leu Gln Phe Thr Tyr Asn His Ile Gln Arg His Val Asn Asp Met Leu  
 20 450 455 460  
 Gly Arg Ile Ala Val Ala Trp Cys Glu Leu Gln Asn His Glu Leu Thr  
 465 470 475 480  
 Leu Trp Asn Glu Ala Arg Lys Leu Asn Pro Asn Ala Ile Ala Ser Ala  
 485 490 495  
 25 Thr Val Gly Arg Arg Val Ser Ala Arg Met Leu Gly Asp Val Met Ala  
 500 505 510  
 Val Ser Thr Cys Val Pro Val Ala Pro Asp Asn Val Ile Val Gln Asn  
 515 520 525  
 Ser Met Arg Val Ser Ser Arg Pro Gly Thr Cys Arg Pro Leu Val Ser  
 30 530 535 540  
 Phe Arg Tyr Glu Asp Gln Gly Pro Leu Ile Glu Gly Gln Leu Gly Glu  
 545 550 555 560  
 Asn Asn Glu Leu Arg Leu Thr Arg Asp Ala Leu Glu Pro Cys Thr Val  
 565 570 575  
 35 Gly His Arg Arg Tyr Phe Ile Phe Gly Gly Gly Tyr Val Tyr Phe Glu  
 580 585 590  
 Glu Tyr Ala Tyr Ser His Gln Leu Ser Arg Ala Asp Val Thr Thr Val  
 595 600 605  
 Ser Thr Phe Ile Asp Leu Asn Ile Thr Met Leu Glu Asp His Glu Phe  
 40 610 615 620  
 Val Pro Leu Glu Val Tyr Thr Arg His Glu Ile Lys Asp Ser Gly Leu  
 625 630 635 640  
 Leu Asp Tyr Thr Glu Val Gln Arg Arg Asn Gln Leu His Asp Leu Arg



		645		650		655										
	Phe	Ala	Asp	Ile	Asp	Thr	Val	Ile	Arg	Ala	Asp	Ala	Asn	Ala	Ala	Met
		660		665		670										
5	Phe	Ala	Gly	Leu	Cys	Ala	Phe	Phe	Glu	Gly	Met	Gly	Asp	Leu	Gly	Arg
		675		680		685										
	Ala	Val	Gly	Lys	Val	Val	Met	Gly	Val	Val	Gly	Gly	Val	Val	Ser	Ala
		690		695		700										
	Val	Ser	Gly	Val	Ser	Ser	Phe	Met	Ser	Asn	Pro	Phe	Gly	Ala	Val	Gly
	705			710		715										
10	Leu	Leu	Val	Leu	Ala	Gly	Leu	Val	Ala	Ala	Phe	Phe	Ala	Phe	Arg	Tyr
		725		730		735										
	Val	Leu	Gln	Leu	Gln	Arg	Asn	Pro	Met	Lys	Ala	Leu	Tyr	Pro	Leu	Thr
		740		745		750										
	Thr	Lys	Glu	Leu	Lys	Thr	Ser	Asp	Pro	Gly	Gly	Val	Gly	Gly	Glu	Gly
15		755		760		765										
	Glu	Glu	Gly	Ala	Glu	Gly	Gly	Gly	Phe	Asp	Glu	Ala	Lys	Leu	Ala	Glu
		770		775		780										
	Ala	Arg	Glu	Met	Ile	Arg	Tyr	Met	Ala	Leu	Val	Ser	Ala	Met	Glu	Arg
	785			790		795										
20	Thr	Glu	His	Lys	Ala	Arg	Lys	Lys	Gly	Thr	Ser	Ala	Leu	Leu	Ser	Ser
		805		810		815										
	Lys	Val	Thr	Asn	Met	Val	Leu	Arg	Lys	Arg	Asn	Lys	Ala	Arg	Tyr	Ser
		820		825		830										
	Pro	Leu	His	Asn	Glu	Asp	Glu	Ala	Gly	Asp	Glu	Asp	Glu	Leu		
25		835		840		845										

## (2) INFORMATION FOR SEQ ID NO:143:

## (i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 20388 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

	GGATCTCCTC	GTTCTCTTGC	GTGATGGACA	CGTCCTCCGC	GGTGGCCGTG	TCGCCTCCCG	60
	GGGCCGTGAG	CTGCTCCTCC	GGGGAGATGG	GGGGGTCTGG	GGTGCCGACA	ACGGCCGGCC	120
40	CGGCCCCGCC	CGAGACCGAG	GACGCCTGGG	GAGTGGGGGT	GCCGCTTTCC	CCCATCCCCA	180
	GGGACAGGTG	GGCCGCCGCC	TCCGTCGCGG	CGGCGGGAGC	CGCGGCCCCC	AGCCGCGCGA	240
	CGTAGCGACA	AAAGTGGCGA	CAGAGGCGCA	TGAGGCGCGC	GCCGTCGGCC	GCGTATCGCG	300
	TGTTTGGCGG	GACGAGCTCG	TCGTAAC TGA	ACAGGAGCAC	GCGGGCGCAG	GTGCCCCACG	360

	GGCCCCACGC	CAGGCGCAGC	GCCGCGACCG	TGTACGGGTC	GTACACGCCT	TGGGCGTCGC	420
	ACGCGACCGG	CAGGGAGACG	AACAGCCCGC	CCGCGCTGGG	GACGCGCGGC	AGGAGGTCCG	480
	GGTGCGCCCG	GATGACGGGG	GCTAGGATCG	CCCCACCGC	ATCCGCCGGC	ACGTAGGCGG	540
	CAAACGCCGA	ACGCCACGGG	GTGCAGTCGC	CGGTGCGGTG	GGCCCCGGTC	TGGGTTTCGA	600
5	CCCGGAAGTT	CGCGGCCGCC	CCGCCGTCGG	GGCGGCCGCG	CACGAGGGCG	GACAGCGGGA	660
	CCCCCGCCGC	CGCCAGGCAC	TCGCTGGAGA	TGATGACGTG	AATCAGCGAG	GCGGGGCTGC	720
	TCGGGTCCCG	GGTGAGATCG	TATTGGACCT	CGTTGGCAAA	GTGCGCGTTC	ATGGCCCCGC	780
	CGGCGGTGCG	AGCCCTTCCC	GGTGCCGGAA	GGGGCGTGGG	TGGGGGGTGC	GTGTGCGCGT	840
	CCTCGGGGCC	CGCGGGCGCA	CGTGCGCTTA	TACGCTGTGT	GTTTCGTCTG	TCCCCAGGGA	900
10	ATCCGGGGCC	AGGACTTTAA	CCTGCTTTTC	GTCGACGAGG	CCAACTTTAT	TCGCCCCGAT	960
	GCGGTCCAGA	CGATTATGGG	CTTTCCTCAAT	CAGGCCAACT	GCAAGATCAT	CTTCGTCTCG	1020
	TCGACCAACA	CCGGGAAGGC	CAGCACGAGC	TTTTTGATACA	ACCTCCGCGG	GGCCGCCGAC	1080
	GAGCTGCTCA	ACGTGGTCAC	CTATATATGC	GACGACCACA	TGCCGCGGGT	GGTGACGCAC	1140
	ACCAACGCCA	CGGCCTGTTC	CTGCTATATC	CTGAACAAAC	CCGTGTTTAT	CACGATGGAC	1200
15	GGCGCCGTTC	GCCGGACGGC	CGATCTGTTT	CTGCCCCACT	CCTTCATGCA	GGAGATCATC	1260
	GGGGGGCAGG	CCCGCGAGAC	CGGCGACGAC	CGCCCCGTCC	TAACAAAGTC	GGCGGGGGAG	1320
	CGGTTTCTGC	TGTACCGCCC	CTCCACCACC	ACCAACAGCG	GCCTGATGGC	CCCCGAGCTG	1380
	TACGTGTACG	TGGACCCGGC	GTTCACGGCC	AACACGCGCG	CCTCCGGCAC	CGGCATCGCG	1440
	GTCGTCGGGA	GGTACCGCGA	CGATTTCATT	ATCTTCGCCC	TGGAGCACTT	TTTCCTCCGC	1500
20	GCGCTCACGG	GATCGGCCCC	CGCGGACATC	GCCCGCTGCG	TCGTGCACAG	CCTCGCCCAG	1560
	GTGCTGGCGC	TGCACCCCGG	GGCGTTTCGC	AGCGTTCGCG	TGGCGGTCGA	GGGCAACAGC	1620
	AGCCAGGACT	CGGCCGTGGC	CATCGCCACA	CACGTGCATA	CCGAGATGCA	CCGCATCCTG	1680
	GCCTCGGCGG	GGGCCAACGG	CCCGGGGCCC	GAGCTCCTCT	TCTATCACTG	CGAGCCGCCC	1740
	GGCGGCGCGG	TATTGTACCC	CTTCTTTCTG	CTCAACAAAC	AGAAGACGCC	CGCCTTCGAA	1800
25	TACTTTATCA	AAAAGTTCAA	CTCCGGGGGC	GTCATGGCGT	CCCAGGAGCT	CGTCTCCGTG	1860
	ACGGTGCGCC	TGCAGACCGA	CCCGGTCGAG	TATCTGTCCG	AGCAGCTCAA	CAACCTCATC	1920
	GAAACCGTCT	CTCCCAACAC	CGACGTCCGC	ATGTACTCCG	GAAAACGCAA	CGGTGCCGCG	1980
	GACGACCTCA	TGGTCGCGGT	CATCATGGCC	ATTTACCTGG	CGGCCCCGAC	CGGGATCCCC	2040
	CCGGCCTTTT	TTCCGATCAC	GCGCACGTCT	TGAGTCTTTC	TTGCCGTTTC	TTTTGTCTCT	2100
30	CTTCTTTTCC	CCCCTCTCTC	CGCAATAAAC	GCCTTCCCGG	AACTGTGTTT	TCCCCCCTA	2160
	CAACAGTGTT	GTCCGTTGGT	TGGGTGGTTG	GGGTGCGGGG	GTGGGCGGGG	GAAGCAAGAA	2220
	AACGGTCGGC	GAACACAACA	TCGGGAAAAC	GGATTCCCGA	ACGTGCGTCT	TCCCAGATTC	2280
	GACACACACC	CCCCTTCTCC	TTAAATAAAC	ACAAACCACA	CGCTCGTTGG	TTGGTTAATG	2340
	CCGGCGCTTT	ATTTACGTCT	TGTTTTTTTG	CGTTTCCTCC	GCGGGTCCCT	TCCAACACG	2400
35	CCTGCCCCCG	CCTCAGGGGT	AGCGGATAAC	CGGGGCCATG	TCGCCGGATT	GCACAACGGC	2460
	GGCGCCGTCG	AACGTACACA	CCCGAACCGC	CGGGGCCAGG	GCCAGGATGT	CCCCGAGTTG	2520
	GCCCCGCTGC	GCCAGCCAGG	CGACCAGCGC	CTCGTAAAGC	GGCAGCCTGC	GCTCGCCGTC	2580
	CTGCATCAGC	ATGGGGGCTT	CGGGGTGGAT	GAGCTGGGCG	GCTTCTCGCG	TGACGCTCTG	2640
	CATCTGCAGG	AGCGCGTTCA	CGTATCCGTC	CTGGGCGCTC	AGCGCGAGCA	GCCGGGGGAT	2700
40	GAGCGTGAGG	ATGAGGGTGG	TTCCTTCGGT	TATGGAGTAG	ACCATGTTGA	GGACGAGCGA	2760
	CCGAGCTCG	GTGTTTACGG	AGGCGAGTTG	CTGGACGTCG	GCCACGAGCG	AGAGACGGGC	2820
	CCCGTTGTAA	TACAGCACGT	TGAGGTCGGG	GAGCTCCCCG	GGCGTCCGGG	GGTCGGGGTT	2880
	GAGGTCCCGG	ATGCCCCGGG	CGACCAGCCG	CGCGACTATC	TCGCGGGCCA	GGGGCGTTGG	2940

	GAGCGGGACC	GGAAACCGCA	GCGTGAGGTC	CAGCGACTCC	AGGCGCACGT	CCGTCGCCTG	3000
	GCCCTCGAAG	ACGGGCGGGA	CGAGGCTGAC	GGGATCCCCG	TTGCAGAGGT	CGACGGGGGA	3060
	GGTGTGCGG	AGATTGACGG	TGCCGGCGTG	CGTGAGCCCC	AGGTCCACGG	GGCAGGCGAC	3120
	GATTCGCGTG	GGCAGCACCC	GCGTGATTAC	CGCGGGGAAG	CGCCTGCGGT	ACGCCAGCAA	3180
5	CAACCCCAAC	GTGTCGGGAC	TAACTCCTCC	GGAGACGAAC	GATTCGTGCG	CCACGTCCGC	3240
	GAGCGCCAGC	TGGCGGCGGA	TGGTCGGCAG	AAAGACCACT	CGACCCTCGC	ACCGCTGCAG	3300
	CGCCGCGGCA	TCGGGGCGCG	AGATACCCGA	GGGGATCGCG	ATGTCTGCTT	CGAAACAATC	3360
	CGTGATCATG	GCGCCGGGCC	GCGAGACACC	GGAACGCGGG	GGTGCGGGAG	GGCCGGAAAG	3420
	CGCAACGCAA	CCGGGACGAT	GATGAAACAG	AGATGGGGGG	CACCGACCGT	GTGGGAGAGG	3480
10	GGGCGGGGCA	GGGCTCAGCA	GCACGCACGG	GGAGGTCTGT	CGTGCGCAGG	AGCCCCAGGT	3540
	GAGAATCAGT	CCCCCGGAGC	TCGGGTCTGG	GTTTTATTGG	GACCTGCCCT	CGGAATCGCG	3600
	GCTCCCAGTC	CAAGCCCCCC	CGGGGGGGCG	GGGACAGGGG	GTGTGTGTGG	GTAAAAGCAA	3660
	CGTCGGAAAA	TCAAACCCAA	TGCCCCAAAC	AGGAAAAAAA	AAAAAGACGG	GCGGGTGGAG	3720
	GGAAAGCTGG	GGAAGAAGAA	GCCAATTTTA	CAGAGACAGG	CCCTTTAGCG	GGGAGGCGTC	3780
15	GTAGATGAGA	TACTGCGTAA	AGTGGGTCTC	TCGCGCGTGG	GCCTCCCCAT	CGCGGGCGCT	3840
	GCGTAGCAGG	GCGGGGTCTC	TGGCGCAGGT	GATCGGGTAG	GCTTCCTGAA	ACAGGCCGCA	3900
	CGGGTCTTCC	ACGAGCTCGC	GGCACCCCGG	CGGGCGCTTA	AACTGCACGT	CGCTGGCAGC	3960
	GGTGGCCGTG	GATACCGCCG	ATCCCGTTTC	CACGATAAGA	CGCTCCAGGC	AGCGATGTTT	4020
	GGCCGTGATG	TCGGCCGCGG	TGAAGAACTT	GAAGCAGGGG	CTGAGGACGG	GCGAGGCCCC	4080
20	GTTGAGGTGA	TAGGCCCCGT	TGTACAGCAG	GTCCCCGTAC	GAGAACCCTG	GCGACGCCCA	4140
	CGGGTTGGCC	GTGGCCGCGA	AGGGCCGCGC	CGGGTCGCTC	TGGCCGTGGT	CGTACATGAG	4200
	GGCTATGACG	TCCCCCTCCT	TGTCCCCCGC	GTACACGCCG	CCGGCCGCGC	GTCCCCGCGG	4260
	GTTGCAGGGC	CGGCGAAAGT	AGTTGATGTC	CGTGGCCACG	GGGGTGGCGA	TGAACTCACA	4320
	CACGGCATCC	TGCCCCGTGGT	CCATGCCCGG	GCGCCGCGGC	ACCTGGGCGC	AGCCAAAGAC	4380
25	CGGGAGGGGC	TGGGCCGGCC	CCAGCCGGTT	TCCCGCCACG	ACCGCGTTGC	GCAGGTACAC	4440
	GGCGGCCGCG	TTGTCTAGCA	GCGGGGGGGC	CCC CGGCGCC	AGGTAAAAGT	TTTGGGGGAG	4500
	GTTGCCCCATG	TCCGTAACGG	GGTTGCGGAC	GGTGGCCGTG	GCCGCGACGG	CGGTGTAGCC	4560
	CACACCCAGG	TCCACGTTTC	CGCGCGGCTG	GGTGAGCGTG	AAGCTGACCC	CCCCGCCCGT	4620
	TTCTGTGGCG	GCCACCTGGA	GCTGGCCCAG	AAAGTACGCC	TCCGACGCGC	GCTCGGAAAA	4680
30	CAGCACGTTC	TCGGTCACGA	AGCGGTCCCTG	CCGCACGACG	GTGAACCCGA	ACCCGGGGTG	4740
	GAGGCCCGTC	TTGAGCTGGT	GATACAGGGC	CACGGGGCTC	ATCTTGAAAGT	ACCCCGCCAT	4800
	GAGCGCGTAG	GTCAGCGCGT	TCTCCCCCGC	CGCGCTCTCG	CGGGCGTGCT	GCACCACGGG	4860
	CTGGCGGATG	GAGGAGAAAGT	AGTTGGCCCC	CAGGGCCGGG	GGGACCAGGG	GGACGTGGCG	4920
	CGCCAGGTCTG	CGCAGGGCCG	GGGGGAAGTT	GGGCGCGTTG	GCCACGTGGT	CGGCGCCCCG	4980
35	AAACAGCGCG	TGGACGGGCA	GGACGTAGAA	GTATTCGCCA	TTTTGGATGG	TGTGGTCCAG	5040
	GTGCTGGGGG	GCCATGAGCA	GCACGCCGGC	GTGCAGCGCC	CCGTCGAAGA	TGCGCATGTT	5100
	GGCCGTCGAC	GCGGTGTTGG	CGCCCGCGTC	GGGCGCCGCG	GAGCACAGCA	GCGCCGTCGT	5160
	GCGCTCGGCC	ATGTTGTGCG	CCAGCACCTG	CAGCGTGAGC	ATGGCGGGCC	CGTCGACGAC	5220
	GACGCGCCCG	TTGTGGAACA	TGGCGTTGAC	CGTGTGGGCC	ACCAGATTGG	CGGGATGCAG	5280
40	CGGGTGGGCG	GGGTCGGTCA	CGGGATCGCT	CGGGCACTCC	TCGCCGGGGG	CGATCTCCGG	5340
	GACCACCATG	TTCTGCAGCG	TGGCGTACAC	GCGGTCTGAAG	CGGACCCCCG	CGGTGCAGCA	5400
	GCGCCCCCGC	GAGAAGGCCG	GCACCAGCAC	GTAATAGTAG	ATTTTGTGGT	GGACGGTCCA	5460
	GTCGGCCGGC	CGGTGCGGCC	GGTCGTCGGC	GGCGTCGGCC	GCGCGGCCTT	GGGTGTTGTG	5520

	CAGCAGCCGG	CCGTCGTTGC	GGTTAAAGTC	GGCCGTCGCC	ACGTTGCACG	CCGCCGCGTA	5580
	GACGGGCTCG	TGTCCCCCG	CGTCAATCCG	GCAGTCTCGG	TGGCGGTCCA	GGGCCGCGTG	5640
	TCGCATAAGG	CCGTCGCAGT	CCCACACGAG	GGGCGGCAGC	AGCGCCGGGT	CGCGCATCAG	5700
	GTGATTACAG	TCGGCCTGAG	CCTGCCCCGC	CAGCTCCGGG	CCCGGCAGGG	TAAAGTCGTC	5760
5	CACCAGCTGG	GCCAGGGCCT	CGACGTGGGC	CACCAGGTCC	CGATACACGG	CCATGCACTC	5820
	CTCGGGGAGG	TCGCCCCCGA	GGTAGGTCAC	GATGTACGAG	ACCAGCGAGT	AGTCGTTTAC	5880
	GAACGCCGCG	CATCGCGTGT	TGTTCCAGTA	GCTGGTGATG	CACTGAGTCA	CGAGCCGCGC	5940
	CAGGCGCAG	AACACGTGCT	CGCTGCCGTG	AATCGCGGCT	TGCAGCAGGT	AAAACACCGC	6000
	CGGGTAGCTG	CGGTCTCTGA	ACGCCCCGCG	GACGGCGGCT	ATGGTAGCCG	GCGCCATGGC	6060
10	GTGGCGGCCA	ACGCCGAGCT	CCAGGCCCCG	GGCGTCACGA	AACGCCACCG	GACACAGCGC	6120
	CAGGGGCAGG	TTGCCGTTGA	CCACGCGCCA	GGTGGCCTGG	ATCGCCCCCG	GACCGGCCGG	6180
	GGGACTTCG	CCGCCGGGAA	GCTCGACGTC	GGCCACGCCC	GCGAAGAAGT	CGAACGCGGG	6240
	GTGCAGCTCC	AGAGCCAGGT	TGGCGTTGTC	GGGCTGCATG	AACTGCTCCG	CGGTCATCTG	6300
	GCACTCGGCG	ACCCACCGGA	CCCGGCCGTG	GGCGAGGCGC	TGCCGCCAGG	CGTTCAGAAA	6360
15	ACGTGCTGCT	ATGTCCGCGC	CGGGGCCGCG	CGGGGCCGCG	ACGTACGCCC	CGTACGGATT	6420
	CGCGGCCTCG	ACGGGGTTCG	GGTTCACGCC	CCCGACGGCC	GCGTCGATGT	TCATGAGCGA	6480
	AGGATGACAC	ACGGTCCCGA	CCGCGTTCTC	CATGGACAGC	CGCAGAACCT	GGTGGTCTCT	6540
	TCCCCAAAAA	AACAGCTGCC	GGGGAGGGAA	CGCGCGGGGC	TCCGGGTGGC	CGGGGGCGGG	6600
	CACCAGGTCC	CCGGCGTGCG	CGGCGAAGCG	CTCCATGGCC	GGGTTGAACA	GCCCCAGGGG	6660
20	CAGGACGAAC	GTCAGGTCCA	TGGCGCCAC	CAGGGGGTAG	GGCACGTTGG	TGGCGGCGTA	6720
	GATGCGCTTC	TCCAGGGCCT	CCAGGAAGAC	CAGCCTGTCT	CCTATGGCCA	CCAGATCCGC	6780
	GCGCACGCGC	GTGTCTGGG	GGGCGCTTTC	GAGTTCATCC	AGCGTCTCCC	GGTTCGCCTC	6840
	GAGTTGCTCC	TCCTGCATCT	CCAGCAGGTG	GCGGCCACG	TCGTCCAGGC	TCCGCACGGC	6900
	CTTGCCCATC	ACCAGCGCCG	TGACGAGGTT	GGCCCCGTTC	AAGACCATCT	CGCCGTAGGT	6960
25	CACCGGCACG	TCGGCCTCGG	TGTCC'TCCAC	CTTCAGGAAG	GACTGCAGGA	GGCGCTGTTT	7020
	GATGGCGGCG	GTGGTGACCA	GCACCCCGTC	GACCGGCCGC	CCGCGCGTGT	CGGCGTGCGT	7080
	CAGGCGGGGC	ACGGCCACGG	ACGGCTGCGT	CGCCGTGGTC	AGGTCCACGA	GCCAGGCCCTC	7140
	GATGGCCTCG	CGGCGATGGC	CCGCC'TTGCC	CAGGAAGAAG	CTCGTGTCGC	AAAAGCTCCG	7200
	CTTCAGCTCG	GCGACCAGGG	TCGCCCGGGC	GACCCTGGTC	GCCAGGCGCC	CGTTGTGCGAG	7260
30	ATATCGTTGC	ATGGGCAACA	GCAGGGCCAG	GGGAGGCGCC	TTCTCCAACA	GCACGTGCAG	7320
	CATCTGGTCG	GCCGTGCCGC	GCTCAAACGC	CCCCAGGACG	GCCTGGACGT	TGCGCGCGAG	7380
	CTGCTGGATG	GCGCGCAGCT	GGCGATGCAG	GCTAATGCCC	GTCCCGTCCA	GGGCCTCCCC	7440
	CGTGAGCAGG	GCAATGGCCT	CGGTGGCCAG	GCTGAAGGCG	GCGTTCAGGG	CCCGGCGGTC	7500
	GATGACCTTC	GTCATGTAAT	TATGCACGGG	CTGCTCGACG	GGGTGCGGGC	CGTCGCGGGC	7560
35	GATGAGGGGC	TGGTGGACCT	CGAACTGCAC	ACGCCCTTCG	TTCATGTAAG	CCAGCTCCGG	7620
	GAAC TTGGTG	CACACGCACG	CCACGGACAG	GCCGAGCTCC	AGAAAGCGCA	CGAGCGACAG	7680
	GGTGTGTCAG	TAGGACCCCA	GCAGGGCGTC	AAACTCTACG	TCATACAGGC	TGTTTTTCGTC	7740
	GGAGCGCACG	CGGGCGAAAA	AATCAAAGAG	TCTGCGGTGG	GACGCCACCT	CGATCGTACT	7800
	CAGGATGGAG	CCGGTGGGCA	CCATGGCCGC	GGCGTACCGG	TAACCCGGGG	GGTCGCGGGC	7860
40	AGGAGCGGCC	ATTGGGTTC	TTGGGGGATT	CGCAGGCTCC	ATCAAGCCAA	GCTCGGGAAG	7920
	GCCAAAGCCCC	TCCCACACAA	CGCCTCACCG	CCGGCGGACG	CGACTAACAA	CCCACGGGCC	7980
	GCCAAAAACC	CCAAGGGGCA	ACCCGACCAA	CAACAGGCGA	GGGGAGGAAA	GGCGTAAAGG	8040
	GGGCGTTGGG	AGGCAAAAAG	AAAGAAAACA	CCCAGACGTA	GGCCCCGAGGA	CCGGCCGGCG	8100

	TCCTCTGTCC	CCGAGCACCC	ACTGTGCCCA	ACAGGCACGG	GGGCGAGCTG	CCCCTGCCTT	8160
	ATATACCCCC	CCGCCACACC	CCCGTTAGAA	CGCGACGGGT	GCCTTCAAGA	TGGCCCTGGT	8220
	CCAAAAGCGT	GCTAGAAAAA	AGTTGGTAAA	GGCGGCAAAG	CAGTCCGCCG	CCGCCACCCA	8280
	CATGGCGGCG	CCGGCCGCGC	AGGCGATTCC	CAGAGAACGG	GCGCGGAGGG	GATCCGTGCG	8340
5	GGGCAGCAGC	TGGCTGGCGG	TGATCCAATG	GAAAAGCCCC	TCGGGACTGA	ACGTCTCATG	8400
	GGCGGCCGCC	ACCAGGGCGC	ACAGGGCCGC	GCCGCCCATG	ATCACGCACA	ACCCCCAAAA	8460
	CACGGGTGGC	GACAACGGCA	GGCGATCCCG	TTTGATGTTT	ACGTACAGGA	GGAGCGCCCC	8520
	TGCCAGCCAC	GTGACATAGT	AGGCGAGGAC	GGCGGCTATA	ATACATGCCG	GCGCCACCGC	8580
	CCGTCCGGTC	CACCCGTAAT	ACATGCCCGC	GGCCACCAGC	TCCAGCGGCT	TGAGGACCAG	8640
10	GAACGACCAA	GCAAACATCA	CCACCCGCTT	GGAAAAGACC	GGCTGGGTGT	GGGGCGGAAG	8700
	ACGCGAGTAG	GCCGAAGTGA	CAAAAAATC	AGACGTGCCG	TACGAGGACA	GCGAAAACCTG	8760
	TTCATCGAGC	GGCAGTTGCG	CGTCTCCCCC	GCCACACGCG	GCCTCGTATA	CCAGCTCGCG	8820
	ATCCAACAAA	GGAACATCAT	CCCGCATTGT	CATGGTCGGT	GCGGGGAGCC	GGCGAGGCAG	8880
	CGAAACCGAA	AGTAGTGCTG	GCGGCGCGGG	CCCGGGTCCG	GACCCAAGCT	TCAGGGATGG	8940
15	GGGGCGGAGG	CCAAAATCAA	ACAAGCACCG	CGCGGGTTCT	ACACACAACC	CCCACCCGGG	9000
	TAGTATCCGC	GGATGCGAGT	GCCTGGCGAA	GTCACGTCCC	AGCAGGATAT	AAACCTCGGC	9060
	CGTTGGGCCC	GGAACCCCCG	AAATTACAC	CCACGCCCTG	ACGCCCAAAT	CATGGGTGGA	9120
	TGTGGTTGCG	GAGCCGCACA	TCCGTGCGTC	CGCCCTCCCC	CGCGGGCTGA	TGACGTGGCG	9180
	GTTAGTCACT	GGGAAGGCAG	GGGGAAGAT	GGGTTGGGGG	AGGAAACGAA	GAAAACACCC	9240
20	AGAGGGCCAC	GTGCGGAATG	CGCCCGGAGT	TGTCTTTAAA	AGGCCGGCCG	TGCGTGACGG	9300
	AAGCCGTGCT	TTGCCCAAGC	ACCGACGCCG	CGATCCACAG	TGGGGGGAGT	TCCTCCGTCC	9360
	GGCCACAACC	CTACGCGCGG	GCGGCACGCG	CGAGAGCAAC	CCACGGGTCC	CGTTCGCGCC	9420
	ACCGCCAGCC	CTTGCTCCCA	CCACCCTCCT	CCCACCACCC	CACTATTCCC	CCCCCCCCAA	9480
	GTCCGCCCCG	TGGCTCGCCG	GCCATGGAGC	TCAGCTATGC	CACCACCCTG	CACCACCGGG	9540
25	ACGTTGTGTT	TTACGTCACG	GCAGACAGAA	ACCGCGCCTA	CTTTGTGTGC	GGGGGGTCCG	9600
	TTTATTCCGT	AGGGCGGCCT	CGGGATTCTC	AGCCGGGGGA	AATTGCCAAG	TTTGGCCTGG	9660
	TGGTCCGGGG	GACAGGCCCC	AAAGACCGCA	TGGTCGCCAA	CTACGTACGA	AGCGAGCTCC	9720
	GCCAGCGCGG	CCTGCGGGAC	GTGCGGCCCG	TGGGGGAGGA	CGAGGTGTTT	CTGGACAGCG	9780
	TGTGTCTGCT	AAACCCGAAC	GTGAGCTCCG	AGCGAGACGT	GATTAATACC	AACGACGTTG	9840
30	AAGTGCTGGA	CGAATGCCTG	GCCGAATACT	GCACCTCGCT	GCGAACCAGC	CCGGGGGTGC	9900
	TGGTGACCGG	GGTGCGCGTG	CGCGCGCGAG	ACAGGGTCAT	CGAGCTATTT	GAGCACCCGG	9960
	CGATCGTCAA	CATTTCCTCG	CGCTTCGCGT	ACACCCCTCT	CCCCTACGTA	TTCGCCCTGG	10020
	CCCAGGCGCA	CCTCCCCCGG	CTCCCGAGCT	CGCTGGAGCC	CCTGGTGAGC	GGCCTGTTTG	10080
	ACGGCATTCC	CGCCCCGCGC	CAGCCCCTGG	ACGCCCGCGA	CCGGCGCACG	GATGTCGTGA	10140
35	TCACGGGCAC	CCGCGCCCCC	AGACCGATGG	CCGGGACCGG	GGCCGGGGGG	GCGGGGGCCA	10200
	AGCGGGCCAC	CGTCAGCGAG	TTCGTGCAAG	TGAAGCACAT	CGACCGTGTT	GTGTCCCCGA	10260
	GCGTCTCTTC	CGCCCCCCCC	CCGAGCGCCC	CCGACGCGAG	TCTGCCGCCC	CCGGGGCTCC	10320
	AGGAGGCCGC	CCCGCCGGGC	CCCCCGCTCA	GGGAGCTGTG	GTGGGTGTTT	TACGCCGGCG	10380
	ACCGGGCGCT	GGAGGAGCCC	CACGCCGAGT	CGGGATTGAC	GCGCGAGGAG	GTCCGCGCCG	10440
40	TGCATGGGTT	CCGGGAGCAG	GCGTGGAAGC	TGTTTGGGTC	GGTGGGGGCT	CCGCGGGCGT	10500
	TTCTCGGGGC	CGCGCTGGCC	CTGAGCCCGA	CCCAAAAGCT	CGCCGTCTAC	TACTATCTCA	10560
	TCCACCGGGA	GCGGCGCATG	TCCCCCTTCC	CCGCGCTCGT	GCGGCTCGTC	GGTCGGTACA	10620
	TCCAGCGCCA	CGGCCTGTAC	GTTCCCGCGC	CCGACGAACC	GACGTTGGCC	GATGCCATGA	10680

	ACGGGCTGTT	CCGCGACGCG	CTGGCGGCCG	GGACCGTGGC	CGAGCAGCTC	CTCATGTTCG	10740
	ACCTCCTCCC	GCCCAAGGAC	GTGCCGGTGG	GGAGCGACGC	GCGGGCCGAC	AGCGCCGCCC	10800
	TGCTGCGCTT	TGTGGACTCG	CAACGCCTGA	CCCCGGGGGG	GTCCGTCTCG	CCCAGACACG	10860
	TCATGTACCT	CGGCGCGTTC	CTGGGCGTGT	TGTACGCCGG	CCACGGACGC	CTGGCCGCGG	10920
5	CCACGCATAC	CGCGCGCCTG	ACGGGCGTGA	CGTCCCTGGT	CCTGACCGTG	GGGGACGTCG	10980
	ACCGGATGTC	CGCGTTTGAC	CGCGGGCCCG	CGGGGGCGGC	TGGCCGCACG	CGAACC GCCG	11040
	GGTACCTGGA	CGCGCTGCTT	ACCGTTTGCC	TGGCTCGCGC	CCAGCACGGC	CAGTCTGTGT	11100
	GAGATATCCC	AATAAAGTGC	AGTCGTTTTT	TAACCCACGG	ATGCCGTTGT	ATGCCTATAC	11160
	GGGGGACTAT	GGGGGGGGGG	GGAAAGGAAA	GGAAACAGGA	ATGGAGAAGG	GAAAGGAACA	11220
10	GAGGCGGTAG	CGGACGCACG	GCGGACACAA	TAACAAACAG	ACCGCGGACA	CGGAGGGAGT	11280
	CGGTTGGGTT	GGGCGTGGAC	GCCGCTGCGT	CCACACACCC	GTTTATTTCG	GTCTCCACAA	11340
	AAATGGGACG	CACGTTTCGA	CCACCCTGAG	GATGCCCGCC	AGGGCCGCGG	TGATCATAAC	11400
	GACCCCGACG	GCGGACGCGG	CCAGAAACCC	GGGGGCGATG	GTGGCGATGG	GCAGCGTGTG	11460
	AAAGGCCAGC	AGATGAATCA	CAGTTCCGTT	GGGGAACAAC	AACAGGGCCA	CGGACGGCAC	11520
15	GTCGCTGGAA	AACACGTTCG	GGGTGCCCGC	CACCGGCCCC	TGGGCCAGCT	GCTGTTGGGT	11580
	GGCATCCGTG	TCCACCAGCA	GCACCGACAT	GACCTCCCCG	GCCGGGGTGT	AGCGCAGAAA	11640
	CACGGCCCCC	ACGAGGCCGA	GGTCGCGCCG	GTTTTTCGGT	CGCACCAGCC	GCTTCGGCTC	11700
	AATCTCCCCG	GCGTGCCCTT	CGCAGGTGGC	GGTGAGATAG	GTGATAAACA	GCGGGCGGGC	11760
	GACGTCAACG	CCCGTAAGCT	TGTATCCGAT	CCCGCGGGGC	AAGGGGGTGT	GGGTGACGAC	11820
20	GTAGCTGGCG	TTGTGGGTGA	TGGGCACGAG	GATCCGGGGC	TCCGCGTTGT	GCGACGGGCC	11880
	GCTACACTGG	TGGGTGGCCT	CCGGGACGAA	GGCGCGGATC	AGGGCGTTGT	AGTGCGCCCA	11940
	GCGCGTGAGA	ACGGAGGCCA	CGCCGCGGGT	CTGTTGTGCC	ATGACGTCCG	CCGGGATGTC	12000
	GGATCGGGTG	GCCATGGCCA	GCGCGTCCAG	GATGAACCCG	CCCTCGGCGA	GATCGAAGCG	12060
	CAGGGAAGCT	GCGCATGGGG	AAAAGTGGTC	CGGGAGCCAG	AAGAGGTTTT	TCTGGTGGTC	12120
25	GGTCTTGGCT	AGCGCGGCCC	GGAGATCGGC	GTGGGTGCGC	GCGGCGACGT	CGGACGTACA	12180
	CAGGGCCGTG	GTTATGAGGA	GGCCCCGGCG	GGCGCGTTCC	CGCTGCTCGG	CCGAGGGCGC	12240
	GCCCCCAGG	AACGGCGCCC	GGAGGACGGC	CGTGGCGTAA	AACAGCGCTC	GGCGGACCAT	12300
	CGGGGCGGTT	AGCGCGCGGC	CGCCGAGAAA	CTCGGCGTAC	AGGGCGTCGA	TCAGGCGGGC	12360
	CGCGCTCGGG	GCCACCGCGC	CATAGGCCGC	GGGGCTGTCC	AACACGAACG	CCAGCTGATA	12420
30	GCCCAGCGCG	TGCGCCGCCA	GGCTCTGCTC	TCGCTCGAGG	ATCGCGGCCA	CCAGATGCCC	12480
	GAGGCGCGCC	TCCAGCCGCA	GGCGGGCCGC	CGGGTCCAAC	ACGGACACGT	TCAGGAACAC	12540
	CGAGTCGGCC	GCGCAGCCCG	CTGCTCCCCG	GGCGGCCAGG	CCGGCCAGCA	CGCGCGAGTG	12600
	GGCCAAAAAG	CCCAGCAGGT	CGGAGAGGCG	AATCGCGTCG	TGGGCGTGGG	CCGCGTTGAC	12660
	GAACGCAAAC	CCCGACGAGG	CGAGCAGCCC	CGCGAGGCGC	CAGAACAGGG	ACGGACGCGC	12720
35	GTCCGTGCGG	GAGCCCGGGT	CCTCCCCCAA	AAACTCCGCA	TAGGCCCGCG	ACATATACTG	12780
	GGCGTAGTTC	GTGCTCTCCT	CGGGGTAGCC	GGCCACCCGC	CGGAGGGCGT	CCAGCGCCGA	12840
	GCCGTTGTCT	GCGGGCGTCG	GGGCCCCCAG	GACAAAGACG	CGATACCTGG	GGCCGGCCCG	12900
	AGGCCCCGGG	AGCACCGCGG	GGGCGTTTTT	GTCGGTCGGA	TTTCCGACCC	GAGCGAGGGT	12960
	CTTGTCGCA	GGCACCCTA	TGATCTCGGC	CGGAGGGCTG	TCCCGCATCG	ATATCACGAG	13020
40	CCCCATGAAG	CCCTTCCCGT	ATCGCGCGCG	CACGAGCGCG	GCGTCGCACC	CGAACGCCAG	13080
	CCCCCCCGTC	GTCCAGACGC	CCACGGGCCA	CGTCGAGGCC	GACGGGGAGA	GGTACACGTA	13140
	CCGACCCGGA	GTCCGTAGCA	GGCCCCCTGG	GGCCAGCCAG	GTCACGGATG	CGTTGTGCAG	13200
	ATGCGCGATG	CTCAGGTTCT	TCGTCGGATG	CCTCGGTGTC	CCCGCGGGCG	GCCCCGGGGG	13260

	CGGCGCGTTG	CGTCGGCCGT	CCGGGTGCCT	CTCGGTCGCC	CCGTCTCTC	CCCGCGGGAA	13320
	CGTAAGCCCC	TCGCGGTCCG	GCGCGGCCGC	GAATGTTACC	CAGGCCCGGG	ACCGCAACAG	13380
	CGCGGAGGCG	CCGGGGTTGT	GCGACAGTCC	CTTGAGCTGG	GTCACCTCGG	CGGGGGGACG	13440
	GGACGTGGGC	CCCGCCTCGG	GGAGCTCGGG	CAGGCTCGCG	TTCCGAGGCC	GGCCGAGCAG	13500
5	ATAGGTCTTT	GGGATGTAAA	GCAGCTGCCC	GGGGTCCCGA	GGAAACTCGG	CCGTGGTGAC	13560
	CAACACGAAA	CAAAAGCGCT	CGGCGTACCA	CCGAAGCATG	GGCACGGATG	CCGTAGTCAG	13620
	GTGAGTTTCG	CCCGGGGGCG	CCAAGCGTCC	GCGCTGGGGG	TCGCTGGCGT	CGGGGGTGTT	13680
	GGGCAACCAC	AGACGCCCCG	TGTTTGTGTC	GCGCCAGTAC	GTGCGGGCCA	ACCCAGACC	13740
	GTGCAAAAAC	CACGGGTCGA	TTTGCTCCGT	CCAGTACGTG	TCATGGCCCC	CGGCAACGCC	13800
10	CACCAGGACC	CCCATCACCA	CCCACAGACC	GGGGCCCATG	GTCGTCCGTC	CCGGCTGCCA	13860
	GTCCGCAGAT	GGGGGGGTGT	CCGTACCCAC	GGCCCAAAGA	GGCTCCGCAC	CTCGGAGGCT	13920
	ATCGGAGGCC	CTTTGTTGCC	GTAAGCGCGG	GCCAAAGGAT	GGGGTGGGGT	GAGGGTAAAA	13980
	GCACAAAGGG	AGTACCAGAC	CGAAAACAAG	GACGGATCGG	CCCGCTCCGT	TTTTCGGTGG	14040
	GGTGCTGATA	CGGTGCCAGC	CCTGGCCCCG	AACCCCCGCG	CTTATGGACA	CACCACACGA	14100
15	CAACAATGCC	TTTTATCTCG	TTCTTTTATT	GCCGTCATCG	CCGGGAGGCC	TTCCGTTCGG	14160
	GCTTCCGTGT	TTGAACTAAA	CTCCCCCAC	CTCGCGGGCA	AACGTGCGCG	CCAGGTGCGG	14220
	TATCTCGGCG	ATGGACCCGG	CGGTTGTGAC	GCGGGTTGGG	ATCATCCCCG	CGGTGAGGCG	14280
	CAACAGGGCG	TCTCGACACC	CGACGGGCGA	CTGATCGTAA	TCCAGGACAA	ATAGATGCAT	14340
	CGGAAGGAGG	CGGTGCGCCA	AGACGTCCAA	GACCCAGGCA	AAAATGTGGT	ACAAGTCCCC	14400
20	GTGCGGGGCC	AGCAGCTCGG	GAACGCGGAA	CAGGGCAAAC	AGCGTGTCCCT	CGATGCGGGG	14460
	CAGAGACCCC	GCGCCGTCCT	CGGGGTCGGG	GCGCGGGGTC	GCCGCGGCGA	CCCCCGTCAG	14520
	CCGGCCCCAG	TCCTCCCGCC	ACCTCCCGCC	GCGCTGCAGG	TACCGCACCG	TGTTGGCGAG	14580
	TAGATCGTAG	ACACGGCGAA	TGGCGGACAG	CATGGCCAGG	TCAAGCCGCT	CGCCCGGGCG	14640
	TTGGCGTCTG	GCCAGGCGGT	CGGCGTGTTC	GGCCTCCGGA	AGGACACCCA	GGACCAGGTT	14700
25	CGTGCCGGGC	GCGGTCGGGG	GCATGAGGGC	CACGAACGCC	AACACGGCCT	GGGGGGTCAT	14760
	GCTTCCCATG	AGGTACCGCG	CGGCCGGGTA	GCACAGCAGG	GAGGCGATAG	GGTGCCGGTC	14820
	GAAACAAGG	GTGAGGGCCG	GGGGCGGGGC	TTGCGGGCCC	ACAGCCTCCC	CCCCGATATG	14880
	AGGAGCCAAA	ACGGCGTCCG	TCGCCGCATA	AGGCGTGCTC	ATTGTTATCT	GGGCGCTGGT	14940
	CATTACCACC	GCCGCCTCCC	CGGCCGATAT	CTCGCCGCGG	TCCAGACGGT	GCTGCGTGTT	15000
30	GTAGATGTTC	GTCAGGTCTT	CGGAGGCCCC	CAGCACCTGC	CAGTAAGTCA	TCGGCTCGGG	15060
	GACGTAGACG	ATATTGTCGC	GCGGCCCCAG	GGCCTCCATC	AGCTGCGCGG	AGGTGGTGGT	15120
	CTTCCCCACC	CCGTGGGGTC	CGTCTATATA	AACCCGCAGC	AGCGTGGGCA	GCTCCGGATC	15180
	CCCGCGGGCT	TCGGAGGCCC	CCTGGCGATG	GCTAGGACGG	GACGCCGCGC	GGCCGTCGGT	15240
	AGGCCCCGTC	GCACGAGCAG	CCTGACCGAA	CGCAGGCGCG	TGCTGTTGGC	CGGCGTGAGA	15300
35	AGCCATACCC	GCTTCTACAA	GGCGTTCGCC	CGAGAGGTGC	GGGAGTTCAA	CGCCACCAGG	15360
	ATTGTGGAA	CGCTGCTGAC	GCTGATGAGC	GGGTCGCTGC	AGGGTCGCTC	GCTGTTTCGAG	15420
	GCCACGCGCG	TCACCTTAAT	ATGCGAAGTG	GACCTCGGGC	CGCGCCGCCC	AGACTGCATC	15480
	TGCGTGTTTC	AATTCGCCAA	TGACAAAACG	TTGGGAGGTG	TGTGCGTCAT	CCTGGAGCTA	15540
	AAGACATGCA	AATCGATTTC	TTCCGGGGAC	ACGGCCAGCA	AACGCGAACA	GCGGACCACG	15600
40	GGCATGAAGC	AGCTGCGCCA	CTCCCTGAAG	CTGCTGCAGT	CGCTCGCGCC	TCCGGGGGAC	15660
	AAGGTGCTCT	ACCTGTGTCC	TATTTTGGTG	TTTGTGCGCG	AGCGTACGCT	GCGCGTCAGC	15720
	CGCGTGACCC	GGCTCGTCCC	GCAAAAGATC	TCCGGCAACA	TCACCGCGGC	CGTGCGGATG	15780
	CTCCAAAGCC	TGTCCACGTA	TGCCGTGCCG	CCGAACCGC	AGACCCGGCG	GTCGCGGCGC	15840

	CGGGTCGCCG	CGACCGCCAG	ACCGCAAAGG	CCCCCTCCC	CGACACGTGA	CCCGGAAGGC	15900
	ACGGCGGGTC	ATCCGGCCCC	ACCAGAGAGC	GACCCCCCT	CCCCAGGGGT	CGTAGGCGTC	15960
	GCTGCGGAGG	GTGGGGGTGT	GCTTCAGAAA	ATCGCGGCGC	TTTTTTGCGT	GCCGGTGGCC	16020
	GCCAAGAGCA	GACCCCGGAC	CAAAACCGAG	TGAGGTTCTG	TGTGTTGTTT	TTTTTCTCG	16080
5	TTTTGTTTTC	TCTTCTTTCC	CCCCCCCCCTC	CCCCGCTTCT	GGCCAAGCAT	CCTCACCTGC	16140
	TTAAGCGGAA	CCCGCGGGCG	CGCGGGGACT	CATTTGTTCG	CGGCGACACC	CACCCGACAA	16200
	CAGCCCCCTG	GTGTAGACCG	CTGTCGCCCC	CGTCTGTTCG	CTCTCCCTTT	TTTCCCCCCC	16260
	TCAAAGAACG	TGGTGTGGG	CGCCGGCCAA	TTCTTCCCAG	AGCGCCGTCG	TCGCCCGCCC	16320
	GCCGCCCTCG	AACATGGACC	CGTACTACCC	TTTCGACGCG	CTGGACGTTT	GGGAACACAG	16380
10	GCGCTTCATC	GTCGCCGACT	CCAGGAGCTT	CATCACCCCC	GAGTTCCCCC	GGGACTTCTG	16440
	GATGTTGCCC	GTGTTCAACA	TCCCCCGGGA	GACGGCGGCG	GAGCGGGCGG	CAGTGCTGCA	16500
	GGCCACGCGC	ACCGCGGCCG	CGGCGGCCCT	GGAGAACGCC	GCCCTCCAGG	CCGCCGAGCT	16560
	GCCCGTCGAC	ATCGAGCGCC	GGATACGCCC	GATCGAGCAG	CAGGTGCATC	ACATCGCCGA	16620
	CGCCCTGGAG	GCGCTGGAGA	CCGCGGCGGC	CGCGGCCGAA	GAGGCGGATG	CCGCGCGGGA	16680
15	CGCCGAGGCG	AGGGGGGAGG	GCGCTGCGGA	CGGGGCGAGC	CCGTCGCCCC	CCGCGGGCCC	16740
	CGCCGCCGCG	GAGATGGAGG	TTGAGATCGT	ACGCAACGAC	CCGCCGCTAC	GATACGATAC	16800
	CAACCTCCCC	GTGGATCTGC	TACACATGGT	GTACGCGGGC	CGCGGGGCCG	CGGGTTCGTC	16860
	GGGAGTCGTC	TTTGGTACCT	GGTACCGCAC	GATCCAGGAA	CGCACCATCG	CGGACTTCCC	16920
	CCTGACCACC	CGCAGCGCCG	ACTTTCGAGA	CGGGCGCATG	TCCAAGACCT	TCATGACCGC	16980
20	GCTGGTCCTG	TCTCTGCAGT	CGTGCGGCCG	GCTGTACGTG	GGCCAGCGCC	ACTATTCCGC	17040
	CTTCGAGTGC	GCCGTGCTGT	GTCTGTATCT	GCTGTACCGA	ACCACCACAG	AGTCTTCCCC	17100
	CGATCGCGAT	CGCGCTCCCG	TTGCGTTCCG	GGACCTGCTG	GCCCGCTGTC	CGCGCTACCT	17160
	GGCGCGTCTG	GCCGCGGTAA	TCGGCGACGA	GAGCGGACGC	CCGCAGTACC	GCTACCGCGA	17220
	CGACAAGCTG	CCCAAAGCGC	AGTTCGCGGC	GGCCGGCGGC	CGCTACGAGC	ACGGGGCCCT	17280
25	GGCCACCCAC	GTGCTGATCG	CCACGTTGGT	GCGCCACGGG	GTGCTACCGG	CGGCCCCGGG	17340
	CGACGTTCCC	CGAGACACCA	GCACCCGCGT	GAACCCCGAC	GACGTGGCCC	ACCGCGACGA	17400
	CGTCAACCGC	GCCGCCGCCG	CGTTTTTGGC	ACGCGGCCAC	AACCTCTTCC	TGTGGGAGGA	17460
	CCAGACGCTG	CTGCGGGCGA	CCGCCAACAC	CATTACGGCC	CTGGCCGTGC	TTCGGCGGCT	17520
	CCTCGCGAAC	GGCAACGTGT	ACGCGGACCG	CCTCGACAAC	CGCCTGCAGC	TGGGCATGCT	17580
30	GATCCCGGGA	GCCGTCCCCG	CGGAGGCCAT	CGCTCGGGGG	GCGTCCGGAT	TGGACTCGGG	17640
	CGCCATAAAA	AGCGGCGACA	ACAACCTGGA	GGCGCTGTGC	GTAACTATG	TACTTCCGCT	17700
	GTATCAGGCA	GACCCACCG	TCGAGCTGAC	CCAGTTGTTT	CCGGGGCTGG	CCGCCCTGTG	17760
	CCTGGACGCC	CAGGCGGGGC	GGCCACTGGC	GTCGACGAGG	CGCGTGGTGG	ATATGTCGTC	17820
	GGGCGCCCGC	CAGGCGGGCG	TCGTGCGCCT	CACCGCGCTG	GAGCTCATCA	ACCGCACCCG	17880
35	CACAAACACC	ACCCCTGTGG	GGGAGATTAT	TAACGCCAC	GATGCCCTTG	GGATACAATA	17940
	CGAACAGGGC	CTGGGGCTGC	TCGCCCAGCA	GGCACGCATC	GGCTTGGCGT	CGAACGCCAA	18000
	GCGATTGCC	ACGTTCAACG	TGGGCAGCGA	CTACGACCTG	TTGTACTTTT	TGTGTCTCGG	18060
	GTTTCATCCC	CAGTACCTGT	CCGTGGCCTA	GGGAAGGGTG	GGGGTGGTGG	TGGTGGGGTG	18120
	TTTTTCTGCT	GTTGTTGTTT	CTGGTCCGCC	TGGTCACAAA	AGGCACGGCG	CCCCGAAACG	18180
40	CGGGCTTTAG	TCCCGGCCCG	GACGTCGGCG	GACACACAAC	AACGGCGGGC	CCCGTGGGTG	18240
	GGTAAGTTGG	TTGCGGGGCA	TCGCTGTATT	CCCTTGCCCC	CTTCCACCCC	CCCTTCCCGT	18300
	TTTGTTTGTT	TGTGCGGGTG	CCCATGGCGT	CGGCGGAAAT	GCGCGAGCGG	TTGGAGGCGC	18360
	CTCTGCCCCG	CCGGCGGGTG	CCCATCTACG	TGGCCGGGTT	TTTGGCCCTG	TACGACAGCG	18420



GGGACCCGGG CGAGCTGGCC CTGGACCCAG ACACGGTGCG TCGGGCCCTG CCTCCGGAGA 18480  
 ACCCCCTGCC GATCAACGTA GACCACCGCG CTCGGTGCGA GGTGGGCCGG GTGCTCGCCG 18540  
 TGGTCAACGA CCCTCGGGGG CCGTTTTTTG TGGGGCTGAT CGCGTGCGTG CAGCTGGAGC 18600  
 GCGTCCTCGA GACGGCCGCC AGCGCCGCTA TTTTGTAGCG CCGCGGACCC GCGCTCTCCC 18660  
 5 GGGAGGAGCG TCTGCTGTAC CTGATCACCA ACTACCTGCC ATCGGTCTCG CTGTCCACAA 18720  
 AACGCCGGGG GGACGAGGTT CCGCCCGACC GCACCCTGTT TGCGCACGTG GCCCTGTGCG 18780  
 CCATCGGGCG GCGCCTTGGA ACCATCGTCA CCTACGACAC CAGCCTAGAC GCGGCCATCG 18840  
 CTCCGTTTCG CCACCTGGAC CCGGCGACGC GCGAGGGGGT GCGACGCGAG GCCGCCGAGG 18900  
 CCGAGCTCGC GCTGGCCGGG CGCACCTGGG CCCCCGGCGT GGAGGCGCTC ACACACACGC 18960  
 10 TGCTCTCCAC CGCCGTCAAC AACATGATGC TCGTGACCG CTGGAGCCTC GTGGCCGAGC 19020  
 GGCGGCGGCA GGCCGGGATC GCCGGACACA CGTACCTTCA GGCGAGCGAA AAATTAAAAA 19080  
 TATGGGGGGC GGAGTCTGCC CTTGCGCCGG AGCGCGGGTA TAAAACCGGC GCCCCGGGTG 19140  
 CCATGGACAC ATCCCCCGCC GCGAGCGTTC CCGCGCCGCA GGTGCGCCGTC CGTGCGCGTC 19200  
 AAGTCGCGTC GTCGTCGTCT TCTTCTTCTT CTTTTCGGC ACCGGCCGAT ATGAACCCCG 19260  
 15 TTTCGGCATC GGGCGCCCCG GCCCTTCCGC CGCCCGGCGA CGGGAGTTAT TTGTGGATCC 19320  
 CCGCCTTTCA TTACAATCAG CTCGTACCG GGCAATCCGC GCCCACCAC CCGCCGCTGA 19380  
 CCGCGTGCGG CCTGCCGGCC GCGGGGACGG TGGCCTACGG ACACCCCGGC GCCGGCCCGT 19440  
 CCCCACACTA CCCGCTCCT CCCGCCACC CGTACCCGGG TATGCTGTTC GCGGGCCCCA 19500  
 GTCCCCTGGA GGCCAGATC GCCGCGCTGG TGGGGGCCAT CGCCGCCGAC CGCCAGGCGG 19560  
 20 GTGGGCTTCC GGCGGCCGCC GGAGACCACG GGATCCGGGG GTCGGCGAAG CGCCGCCGAC 19620  
 ACGAGGTGGA GCAGCCGGAG TACGACTGCG GCCGTGACGA GCCGACCGG GACTTCCCGT 19680  
 ATTACCCGGG CGAGGCCCGC CCCGAGCCGC GCCCGGTGCA CTCCCGGCGC GCCGCGCGCC 19740  
 AGGCTTCCGG GCCCCACGAA ACCATCACGG CGCTGGTGGG GGCGGTGACG TCCCTGCAGC 19800  
 AGGAACTGGC GCACATGCGC GCGCGTACCC ACGCCCCCTA CGGGCCGTAT CCGCCGGTGG 19860  
 25 GGCCCTACCA CCACCCACAC GCAGACACGG AGACCCCGC CCAACCACCC CGTACCCCG 19920  
 CCGAGGCCGT CTATCTGCCG CCGCCGACA TCGCCCCCCC GGGGCCCTCT CTATCCGGGG 19980  
 CGGTCCCCC ACCCTCGTAT CCCCAGTTG CGGTTACCCC CGGTCCCGCC CCCCCTAC 20040  
 ATCAGCCCTC CCCCACACAC GCCACCCCG CTCGCGCGCC GCCGGGACCC ACGCTCCCC 20100  
 CCGCCGCGAG CTTACCCCAA CCGAGGCGC CCGGCGCGGA GGCCGGCGCC TTAGTTAACG 20160  
 30 CCAGCAGCGC GGCCACGTG AACGTGGACA CGGCCCGGGC CGCCGATCTG TTTGTGTCAC 20220  
 AGATGATGGG GTCCCGCTAA CTCGCTCCA GGATCCGGAC TTGGGGGGG TGTGTGTTTT 20280  
 CATATATTTT AAATAAACAA ACAACCGGAC AAAAGTATAC CCACTTCGTG TGCTTGTGTT 20340  
 TTTGTTTGAG AGGGGGGGG TGGAGTGGG GGGAAAGTG GCCGAAT 20388

35 (2) INFORMATION FOR SEQ ID NO:144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 262 amino acids  
 (B) TYPE: amino acid  
 40 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

5 Met Asn Ala His Phe Ala Asn Glu Val Gln Tyr Asp Leu Thr Arg Asp  
 1 5 10 15  
 Pro Ser Ser Pro Ala Ser Leu Ile His Val Ile Ile Ser Ser Glu Cys  
 20 25 30  
 Leu Ala Ala Ala Gly Val Pro Leu Ser Ala Leu Val Arg Gly Arg Pro  
 35 40 45  
 10 Asp Gly Gly Ala Ala Ala Asn Phe Arg Val Glu Thr Gln Thr Arg Ala  
 50 55 60  
 His Ala Thr Gly Asp Cys Thr Pro Trp Arg Ser Ala Phe Ala Ala Tyr  
 65 70 75 80  
 Val Pro Ala Asp Ala Val Gly Ala Ile Leu Ala Pro Val Ile Pro Ala  
 15 85 90 95  
 His Pro Asp Leu Leu Pro Arg Val Pro Ser Ala Gly Gly Leu Phe Val  
 100 105 110  
 Ser Leu Pro Val Ala Cys Asp Ala Gln Gly Val Tyr Asp Pro Tyr Thr  
 115 120 125  
 20 Val Ala Ala Leu Arg Leu Ala Trp Gly Pro Trp Ala Thr Cys Ala Arg  
 130 135 140  
 Val Leu Leu Phe Ser Tyr Asp Glu Leu Val Pro Pro Asn Thr Arg Tyr  
 145 150 155 160  
 Ala Ala Asp Gly Ala Arg Leu Met Arg Leu Cys Arg His Phe Cys Arg  
 25 165 170 175  
 Tyr Val Ala Arg Leu Gly Ala Ala Ala Pro Ala Ala Ala Thr Glu Ala  
 180 185 190  
 Ala Ala His Leu Ser Leu Gly Met Gly Glu Ser Gly Thr Pro Thr Pro  
 195 200 205  
 30 Gln Ala Ser Ser Val Ser Gly Gly Ala Gly Pro Ala Val Val Gly Thr  
 210 215 220  
 Pro Asp Pro Pro Ile Ser Pro Glu Glu Gln Leu Thr Ala Pro Gly Gly  
 225 230 235 240  
 Asp Thr Ala Thr Ala Glu Asp Val Ser Ile Thr Gln Glu Asn Glu Glu  
 35 245 250 255  
 Ile Xaa Xaa Xaa Xaa Xaa  
 260

## (2) INFORMATION FOR SEQ ID NO:145:

40

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 423 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

	Val	Pro	Glu	Gly	Ala	Trp	Val	Gly	Gly	Ala	Cys	Ala	Arg	Pro	Arg	Gly	
	1				5					10					15		
10	Pro	Arg	Ala	His	Val	Arg	Leu	Tyr	Ala	Val	Cys	Phe	Val	Cys	Pro	Gln	
				20					25					30			
	Gly	Ile	Arg	Gly	Gln	Asp	Phe	Asn	Leu	Leu	Phe	Val	Asp	Glu	Ala	Asn	
		35				40						45					
	Phe	Ile	Arg	Pro	Asp	Ala	Val	Gln	Thr	Ile	Met	Gly	Phe	Leu	Asn	Gln	
15		50				55					60						
	Ala	Asn	Cys	Lys	Ile	Ile	Phe	Val	Ser	Ser	Thr	Asn	Thr	Gly	Lys	Ala	
	65				70					75					80		
	Ser	Thr	Ser	Phe	Leu	Tyr	Asn	Leu	Arg	Gly	Ala	Ala	Asp	Glu	Leu	Leu	
				85					90					95			
20	Asn	Val	Val	Thr	Tyr	Ile	Cys	Asp	Asp	His	Met	Pro	Arg	Val	Val	Thr	
				100					105					110			
	His	Thr	Asn	Ala	Thr	Ala	Cys	Ser	Cys	Tyr	Ile	Leu	Asn	Lys	Pro	Val	
		115				120							125				
	Phe	Ile	Thr	Met	Asp	Gly	Ala	Val	Arg	Arg	Thr	Ala	Asp	Leu	Phe	Leu	
25		130				135					140						
	Pro	Asp	Ser	Phe	Met	Gln	Glu	Ile	Ile	Gly	Gly	Gln	Ala	Arg	Glu	Thr	
	145				150					155				160			
	Gly	Asp	Asp	Arg	Pro	Val	Leu	Thr	Lys	Ser	Ala	Gly	Glu	Arg	Phe	Leu	
				165					170					175			
30	Leu	Tyr	Arg	Pro	Ser	Thr	Thr	Thr	Asn	Ser	Gly	Leu	Met	Ala	Pro	Glu	
				180					185					190			
	Leu	Tyr	Val	Tyr	Val	Asp	Pro	Ala	Phe	Thr	Ala	Asn	Thr	Arg	Ala	Ser	
		195				200							205				
	Gly	Thr	Gly	Ile	Ala	Val	Val	Gly	Arg	Tyr	Arg	Asp	Asp	Phe	Ile	Ile	
35		210				215					220						
	Phe	Ala	Leu	Glu	His	Phe	Phe	Leu	Arg	Ala	Leu	Thr	Gly	Ser	Ala	Pro	
	225				230					235				240			
	Ala	Asp	Ile	Ala	Arg	Cys	Val	Val	His	Ser	Leu	Ala	Gln	Val	Leu	Ala	
				245					250				255				
40	Leu	His	Pro	Gly	Ala	Phe	Arg	Ser	Val	Arg	Val	Ala	Val	Glu	Gly	Asn	
				260					265				270				
	Ser	Ser	Gln	Asp	Ser	Ala	Val	Ala	Ile	Ala	Thr	His	Val	His	Thr	Glu	
		275				280						285					

420

Met His Arg Ile Leu Ala Ser Ala Gly Ala Asn Gly Pro Gly Pro Glu  
 290 295 300  
 Leu Leu Phe Tyr His Cys Glu Pro Pro Gly Gly Ala Val Leu Tyr Pro  
 305 310 315 320  
 5 Phe Phe Leu Leu Asn Lys Gln Lys Thr Pro Ala Phe Glu Tyr Phe Ile  
 325 330 335  
 Lys Lys Phe Asn Ser Gly Gly Val Met Ala Ser Gln Glu Leu Val Ser  
 340 345 350  
 Val Thr Val Arg Leu Gln Thr Asp Pro Val Glu Tyr Leu Ser Glu Gln  
 10 355 360 365  
 Leu Asn Asn Leu Ile Glu Thr Val Ser Pro Asn Thr Asp Val Arg Met  
 370 375 380  
 Tyr Ser Gly Lys Arg Asn Gly Ala Ala Asp Asp Leu Met Val Ala Val  
 385 390 395 400  
 15 Ile Met Ala Ile Tyr Leu Ala Ala Pro Thr Gly Ile Pro Pro Ala Phe  
 405 410 415  
 Phe Pro Ile Thr Arg Thr Ser  
 420

20 (2) INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 355 amino acids  
 (B) TYPE: amino acid  
 25 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

Val Leu Leu Ser Pro Ala Pro Pro Pro Leu Pro His Gly Arg Cys Pro  
 1 5 10 15  
 Pro Ser Leu Phe His His Arg Pro Gly Cys Val Ser Gly Pro Pro Ala  
 20 25 30  
 35 Pro Pro Arg Ser Gly Val Ser Arg Pro Gly Ala Met Ile Thr Asp Cys  
 35 40 45  
 Phe Glu Ala Asp Ile Ala Ile Pro Ser Gly Ile Ser Arg Pro Asp Ala  
 50 55 60  
 40 Ala Ala Leu Gln Arg Cys Glu Gly Arg Val Val Phe Leu Pro Thr Ile  
 65 70 75 80  
 Arg Arg Gln Leu Ala Asp Val Ala His Glu Ser Phe Val Ser Gly Gly  
 85 90 95

Val Ser Pro Asp Thr Leu Gly Leu Leu Leu Ala Tyr Arg Arg Arg Phe  
                     100                    105                    110  
 Pro Ala Val Ile Thr Arg Val Leu Pro Thr Arg Ile Val Ala Cys Pro  
                     115                    120                    125  
 5 Val Asp Leu Gly Leu Thr His Ala Gly Thr Val Asn Leu Arg Asn Thr  
                     130                    135                    140  
 Ser Pro Val Asp Leu Cys Asn Gly Asp Pro Val Ser Leu Val Pro Pro  
 145                    150                    155                    160  
 Val Phe Glu Gly Gln Ala Thr Asp Val Arg Leu Glu Ser Leu Asp Leu  
 10                    165                    170                    175  
 Thr Leu Arg Phe Pro Val Pro Leu Pro Thr Pro Leu Ala Arg Glu Ile  
                     180                    185                    190  
 Val Ala Arg Leu Val Arg Ile Arg Asp Leu Asn Pro Asp Pro Arg Thr  
                     195                    200                    205  
 15 Pro Gly Glu Leu Pro Asp Leu Asn Val Leu Tyr Tyr Asn Gly Ala Arg  
                     210                    215                    220  
 Leu Ser Leu Val Ala Asp Val Gln Gln Leu Ala Ser Val Asn Thr Glu  
 225                    230                    235                    240  
 Leu Arg Ser Leu Val Leu Asn Met Val Tyr Ser Ile Thr Glu Gly Thr  
 20                    245                    250                    255  
 Thr Leu Ile Leu Thr Leu Ile Pro Arg Leu Leu Ala Leu Ser Ala Gln  
                     260                    265                    270  
 Asp Gly Tyr Val Asn Ala Leu Leu Gln Met Gln Ser Val Thr Arg Glu  
                     275                    280                    285  
 25 Ala Ala Gln Leu Ile His Pro Glu Ala Pro Met Leu Met Gln Asp Gly  
                     290                    295                    300  
 Glu Arg Arg Leu Pro Leu Tyr Glu Ala Leu Val Ala Trp Leu Ala His  
 305                    310                    315                    320  
 Ala Gly Gln Leu Gly Asp Ile Leu Ala Pro Ala Val Arg Val Cys Thr  
 30                    325                    330                    335  
 Phe Asp Gly Ala Ala Val Val Gln Ser Gly Asp Met Ala Pro Val Ile  
                     340                    345                    350  
 Arg Tyr Pro  
                     355  
 35

## (2) INFORMATION FOR SEQ ID NO:147:

## (i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 1382 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

5 Val Trp Glu Gly Leu Gly Leu Pro Glu Leu Gly Leu Met Glu Pro Ala  
 1 5 10 15  
 Asn Pro Pro Arg Asn Pro Met Ala Ala Pro Ala Arg Asp Pro Pro Gly  
 20 25 30  
 Tyr Arg Tyr Ala Ala Ala Met Val Pro Thr Gly Ser Ile Leu Ser Thr  
 10 35 40 45  
 Ile Glu Val Ala Ser His Arg Arg Leu Phe Asp Phe Phe Ala Arg Val  
 50 55 60  
 Arg Ser Asp Glu Asn Ser Leu Tyr Asp Val Glu Phe Asp Ala Leu Leu  
 65 70 75 80  
 15 Gly Ser Tyr Cys Asn Thr Leu Ser Leu Val Arg Phe Leu Glu Leu Gly  
 85 90 95  
 Leu Ser Val Ala Cys Val Cys Thr Lys Phe Pro Glu Leu Ala Tyr Met  
 100 105 110  
 Asn Glu Gly Arg Val Gln Phe Glu Val His Gln Pro Leu Ile Ala Arg  
 115 120 125  
 20 Asp Gly Pro His Pro Val Glu Gln Pro Val His Asn Tyr Met Thr Lys  
 130 135 140  
 Val Ile Asp Arg Arg Ala Leu Asn Ala Ala Phe Ser Leu Ala Thr Glu  
 145 150 155 160  
 25 Ala Ile Ala Leu Leu Thr Gly Glu Ala Leu Asp Gly Thr Gly Ile Ser  
 165 170 175  
 Leu His Arg Gln Leu Arg Ala Ile Gln Gln Leu Ala Arg Asn Val Gln  
 180 185 190  
 Ala Val Leu Gly Ala Phe Glu Arg Gly Thr Ala Asp Gln Met Leu His  
 195 200 205  
 30 Val Leu Leu Glu Lys Ala Pro Pro Leu Ala Leu Leu Leu Pro Met Gln  
 210 215 220  
 Arg Tyr Leu Asp Asn Gly Arg Leu Ala Thr Arg Val Ala Arg Ala Thr  
 225 230 235 240  
 35 Leu Val Ala Glu Leu Lys Arg Ser Phe Cys Asp Thr Ser Phe Phe Leu  
 245 250 255  
 Gly Lys Ala Gly His Arg Arg Glu Ala Ile Glu Ala Trp Leu Val Asp  
 260 265 270  
 Leu Thr Thr Ala Thr Gln Pro Ser Val Ala Val Pro Arg Leu Thr His  
 275 280 285  
 40 Ala Asp Thr Arg Gly Arg Pro Val Asp Gly Val Leu Val Thr Thr Ala  
 290 295 300  
 Ala Ile Lys Gln Arg Leu Leu Gln Ser Phe Leu Lys Val Glu Asp Thr  
 423

	305		310		315		320
	Glu	Ala	Asp	Val	Pro	Val	Thr
			325			330	
	Asn	Leu	Val	Thr	Ala	Leu	Val
5			340			345	
	Asp	Val	Gly	Arg	His	Leu	Leu
			355			360	
	Asn	Arg	Glu	Thr	Leu	Asp	Glu
			370			375	
10	Val	Arg	Ala	Asp	Leu	Val	Ala
			385			390	
	Ala	Leu	Glu	Lys	Arg	Ile	Tyr
			405			410	
	Val	Gly	Ala	Met	Asp	Leu	Thr
15			420			425	
	Pro	Ala	Met	Glu	Arg	Phe	Ala
			435			440	
	Pro	Gly	His	Pro	Glu	Pro	Arg
			450			455	
20	Trp	Gly	Lys	Asp	His	Gln	Val
			465			470	
	Gly	Thr	Val	Cys	His	Pro	Ser
			485			490	
	Gly	Val	Asn	His	Asp	Pro	Val
25			500			505	
	Val	Ala	Ala	Pro	Ala	Gly	Pro
			515			520	
	Asn	Ala	Trp	Arg	Gln	Arg	Leu
			530			535	
30	Glu	Cys	Gln	Met	Thr	Ala	Glu
			545			550	
	Leu	Ala	Leu	Glu	Leu	His	Pro
			565			570	
	Asp	Val	Glu	Leu	Pro	Gly	Gly
35			580			585	
	Ile	Gln	Ala	Thr	Trp	Arg	Val
			595			600	
	Cys	Pro	Val	Ala	Phe	Arg	Asp
			610			615	
40	Ala	Met	Ala	Pro	Ala	Thr	Ile
			625			630	
	Arg	Ser	Tyr	Pro	Ala	Val	Phe
			645			650	

Ser Glu His Val Phe Cys Ala Arg Leu Val Thr Gln Cys Ile Thr Ser  
 660 665 670  
 Tyr Trp Asn Asn Thr Arg Cys-Ala Ala Phe Val Asn Asp Tyr Ser Leu  
 675 680 685  
 5 Val Ser Tyr Ile Val Thr Tyr Leu Gly Gly Asp Leu Pro Glu Glu Cys  
 690 695 700  
 Met Ala Val Tyr Arg Asp Leu Val Ala His Val Glu Ala Gln Leu Val  
 705 710 715 720  
 Asp Asp Phe Thr Leu Pro Gly Pro Glu Leu Gly Gly Gln Ala Gln Ala  
 10 725 730 735  
 Glu Leu Asn His Leu Met Arg Asp Pro Ala Leu Leu Pro Pro Leu Val  
 740 745 750  
 Trp Asp Cys Asp Gly Leu Met Arg His Ala Ala Leu Asp Arg His Arg  
 755 760 765  
 15 Asp Cys Arg Ile Asp Ala Gly Gly His Glu Pro Val Tyr Ala Ala Ala  
 770 775 780  
 Cys Asn Val Ala Thr Ala Asp Phe Asn Arg Asn Asp Gly Arg Leu Leu  
 785 790 795 800  
 His Asn Thr Gln Ala Arg Ala Ala Asp Ala Ala Asp Asp Arg Pro His  
 20 805 810 815  
 Arg Pro Ala Asp Trp Thr Val His His Lys Ile Tyr Tyr Tyr Val Leu  
 820 825 830  
 Val Pro Ala Phe Ser Arg Gly Arg Cys Cys Thr Ala Gly Val Arg Phe  
 835 840 845  
 25 Asp Arg Val Tyr Ala Thr Leu Gln Asn Met Val Val Pro Glu Ile Ala  
 850 855 860  
 Pro Gly Glu Glu Cys Pro Ser Asp Pro Val Thr Asp Pro Ala His Pro  
 865 870 875 880  
 Leu His Pro Ala Asn Leu Val Ala Asn Thr Val Asn Ala Met Phe His  
 30 885 890 895  
 Asn Gly Arg Val Val Val Asp Gly Pro Ala Met Leu Thr Leu Gln Val  
 900 905 910  
 Leu Ala His Asn Met Ala Glu Arg Thr Thr Ala Leu Leu Cys Ser Ala  
 915 920 925  
 35 Ala Pro Asp Ala Gly Ala Asn Thr Ala Ser Thr Ala Asn Met Arg Ile  
 930 935 940  
 Phe Asp Gly Ala Leu His Ala Gly Val Leu Leu Met Ala Pro Gln His  
 945 950 955 960  
 Leu Asp His Thr Ile Gln Asn Gly Glu Tyr Phe Tyr Val Leu Pro Val  
 40 965 970 975  
 His Ala Leu Phe Ala Gly Ala Asp His Val Ala Asn Ala Pro Asn Phe  
 980 985 990  
 Pro Pro Ala Leu Arg Asp Leu Ala Arg His Val Pro Leu Val Pro Pro



	995	1000	1005
	Ala Leu Gly Ala Asn Tyr Phe Ser Ser Ile Arg Gln Pro Val Val Gln		
	1010	1015	1020
	His Ala Arg Glu Ser Ala Ala Gly Glu Asn Ala Leu Thr Tyr Ala Leu		
5	1025	1030	1035 104
	Met Ala Gly Tyr Phe Lys Met Ser Pro Val Tyr His Gln Leu Lys Thr		
	1045	1050	1055
	Gly Leu His Pro Gly Phe Gly Phe Thr Val Val Arg Gln Asp Arg Phe		
	1060	1065	1070
10	Val Thr Glu Asn Val Leu Phe Ser Ala Ser Glu Ala Tyr Phe Leu Gly		
	1075	1080	1085
	Gln Leu Gln Val Ala Arg His Glu Thr Gly Gly Gly Val Ser Phe Thr		
	1090	1095	1100
	Leu Thr Gln Pro Arg Gly Asn Val Asp Leu Gly Val Gly Tyr Thr Ala		
15	1105	1110	1115 112
	Val Ala Ala Thr Ala Thr Val Arg Asn Pro Val Thr Asp Met Gly Asn		
	1125	1130	1135
	Leu Pro Gln Asn Phe Tyr Leu Gly Arg Gly Ala Pro Pro Leu Leu Asp		
	1140	1145	1150
20	Asn Ala Ala Ala Val Tyr Leu Arg Asn Ala Val Val Ala Gly Asn Arg		
	1155	1160	1165
	Leu Gly Pro Ala Gln Pro Leu Pro Val Phe Gly Cys Ala Gln Val Pro		
	1170	1175	1180
	Arg Arg Ala Gly Met Asp His Gly Gln Asp Ala Val Cys Glu Phe Ile		
25	1185	1190	1195 120
	Ala Thr Pro Val Ala Thr Asp Ile Asn Tyr Phe Arg Arg Pro Cys Asn		
	1205	1210	1215
	Pro Arg Gly Arg Ala Ala Gly Gly Val Tyr Ala Gly Asp Lys Glu Gly		
	1220	1225	1230
30	Asp Val Ile Ala Leu Met Tyr Asp His Gly Gln Ser Asp Pro Ala Arg		
	1235	1240	1245
	Pro Phe Ala Ala Thr Ala Asn Pro Trp Ala Ser Gln Arg Phe Ser Tyr		
	1250	1255	1260
	Gly Asp Leu Leu Tyr Asn Gly Ala Tyr His Leu Asn Gly Asp Val Leu		
35	1265	1270	1275 128
	Ser Pro Cys Phe Lys Phe Phe Thr Ala Ala Asp Ile Thr Ala Lys His		
	1285	1290	1295
	Arg Cys Leu Glu Arg Leu Ile Val Glu Thr Gly Ser Ala Val Ser Thr		
	1300	1305	1310
40	Ala Thr Ala Ala Ser Asp Val Gln Phe Lys Arg Pro Pro Gly Cys Arg		
	1315	1320	1325
	Glu Leu Val Glu Asp Pro Cys Gly Leu Phe Gln Glu Ala Tyr Pro Ile		
	1330	1335	1340

Thr Cys Ala Ser Asp Pro Ala Leu Leu Arg Ser Ala Arg Asp Gly Glu  
 1345                      1350                      1355                      136  
 Ala His Ala Arg Glu Thr His Phe Thr Gln Tyr Leu Ile Tyr Asp Asp  
                          1365                      1370                      1375  
 5    Leu Lys Gly Leu Ser Leu  
                          1380

## (2) INFORMATION FOR SEQ ID NO:148:

## 10            (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 222 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

20    Met Thr Met Arg Asp Asp Val Pro Leu Leu Asp Arg Glu Leu Val Tyr  
       1                      5                      10                      15  
       Glu Ala Ala Cys Gly Gly Glu Asp Gly Glu Leu Pro Leu Asp Glu Gln  
                          20                      25                      30  
       Phe Ser Leu Ser Ser Tyr Gly Thr Ser Asp Phe Phe Val Ser Ser Ala  
 25                      35                      40                      45  
       Tyr Ser Arg Leu Pro Pro His Thr Gln Pro Val Phe Ser Lys Arg Val  
                          50                      55                      60  
       Val Met Phe Ala Trp Ser Phe Leu Val Leu Lys Pro Leu Glu Leu Val  
       65                      70                      75                      80  
 30    Ala Ala Gly Met Tyr Tyr Gly Trp Thr Gly Arg Ala Val Ala Pro Ala  
                          85                      90                      95  
       Cys Ile Ile Ala Ala Val Leu Ala Tyr Tyr Val Thr Trp Leu Ala Arg  
                          100                      105                      110  
       Ala Leu Leu Leu Tyr Val Asn Ile Lys Arg Asp Arg Leu Pro Leu Ser  
 35                      115                      120                      125  
       Pro Pro Val Phe Trp Gly Leu Cys Val Ile Met Gly Gly Ala Ala Leu  
                          130                      135                      140  
       Cys Ala Leu Val Ala Ala Ala His Glu Thr Phe Ser Pro Asp Gly Leu  
       145                      150                      155                      160  
 40    Phe His Trp Ile Thr Ala Ser Gln Leu Leu Pro Arg Thr Asp Pro Leu  
                          165                      170                      175  
       Arg Ala Arg Ser Leu Gly Ile Ala Cys Ala Ala Gly Ala Ala Met Trp  
                          180                      185                      190

Val Ala Ala Ala Asp Cys Phe Ala Ala Phe Thr Asn Phe Phe Leu Ala  
 195 200 205  
 Arg Phe Trp Thr Arg Ala Ile Leu Lys Ala Pro Val Ala Phe  
 210 215 220

5

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 627 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

Val Gly Arg Gln Gly Glu Arg Trp Val Gly Gly Gly Asn Glu Glu Asn  
 1 5 10 15  
 20 Thr Gln Arg Ala Thr Ser Gly Met Arg Pro Glu Leu Ser Leu Lys Gly  
 20 25 30  
 Arg Pro Cys Val Thr Glu Ala Val Val Cys Pro Ser Thr Asp Ala Ala  
 35 40 45  
 Ile His Ser Gly Gly Ser Ser Val Arg Pro Gln Pro Tyr Ala Arg  
 25 50 55 60  
 Ala Ala Arg Ala Arg Ala Thr His Gly Ser Arg Ser Arg His Arg Gln  
 65 70 75 80  
 Pro Leu Leu Pro Pro Pro Ser Ser His His Pro Thr Ile Pro Pro Pro  
 85 90 95  
 30 Pro Ser Pro Pro Arg Gly Ser Pro Ala Met Glu Leu Ser Tyr Ala Thr  
 100 105 110  
 Thr Leu His His Arg Asp Val Val Phe Tyr Val Thr Ala Asp Arg Asn  
 115 120 125  
 Arg Ala Tyr Phe Val Cys Gly Gly Ser Val Tyr Ser Val Gly Arg Pro  
 35 130 135 140  
 Arg Asp Ser Gln Pro Gly Glu Ile Ala Lys Phe Gly Leu Val Val Arg  
 145 150 155 160  
 Gly Thr Gly Pro Lys Asp Arg Met Val Ala Asn Tyr Val Arg Ser Glu  
 165 170 175  
 40 Leu Arg Gln Arg Gly Leu Arg Asp Val Arg Pro Val Gly Glu Asp Glu  
 180 185 190  
 Val Phe Leu Asp Ser Val Cys Leu Leu Asn Pro Asn Val Ser Ser Asp  
 195 200 205

Val Ile Asn Thr Asn Asp Val Glu Val Leu Asp Glu Cys Leu Ala Glu  
 210 215 220  
 Tyr Cys Thr Ser Leu Arg Thr Ser Pro Gly Val Leu Val Thr Gly Val  
 225 230 235 240  
 5 Arg Val Arg Ala Arg Asp Arg Val Ile Glu Leu Phe Glu His Pro Ala  
 245 250 255  
 Ile Val Asn Ile Ser Ser Arg Phe Ala Tyr Thr Pro Ser Pro Tyr Val  
 260 265 270  
 Phe Ala Gln Ala His Leu Pro Arg Leu Pro Ser Ser Leu Glu Pro Leu  
 10 275 280 285  
 Val Ser Gly Leu Phe Asp Gly Ile Pro Ala Pro Arg Gln Pro Leu Asp  
 290 295 300  
 Ala Arg Asp Arg Arg Thr Asp Val Val Ile Thr Gly Thr Arg Ala Pro  
 305 310 315 320  
 15 Arg Pro Met Ala Gly Thr Gly Ala Gly Gly Ala Gly Ala Lys Arg Ala  
 325 330 335  
 Thr Val Ser Glu Phe Val Gln Val Lys His Ile Asp Arg Val Val Ser  
 340 345 350  
 Pro Ser Val Ser Ser Ala Pro Pro Pro Ser Ala Pro Asp Ala Ser Leu  
 20 355 360 365  
 Pro Pro Pro Gly Leu Gln Glu Ala Ala Pro Pro Gly Pro Pro Leu Arg  
 370 375 380  
 Glu Leu Trp Trp Val Phe Tyr Ala Gly Asp Arg Ala Leu Glu Glu Pro  
 385 390 395 400  
 25 His Ala Glu Ser Gly Leu Thr Arg Glu Glu Val Arg Ala Val His Gly  
 405 410 415  
 Phe Arg Glu Gln Ala Trp Lys Leu Phe Gly Ser Val Gly Ala Pro Arg  
 420 425 430  
 Ala Phe Leu Gly Ala Ala Leu Ser Pro Thr Gln Lys Leu Ala Val Tyr  
 30 435 440 445  
 Tyr Tyr Leu Ile His Arg Glu Arg Arg Met Ser Pro Phe Pro Ala Leu  
 450 455 460  
 Val Arg Leu Val Gly Arg Tyr Ile Gln Arg His Gly Val Pro Ala Pro  
 465 470 475 480  
 35 Asp Glu Pro Thr Leu Ala Asp Ala Met Asn Gly Leu Phe Arg Asp Ala  
 485 490 495  
 Ala Gly Thr Val Ala Glu Gln Leu Leu Met Phe Asp Leu Leu Pro Pro  
 500 505 510  
 Lys Asp Val Pro Val Gly Ser Asp Ala Arg Ala Asp Ser Ala Ala Leu  
 40 515 520 525  
 Leu Arg Phe Val Asp Ser Gln Arg Leu Thr Pro Gly Gly Ser Val Ser  
 530 535 540  
 Pro Glu His Val Met Tyr Leu Gly Ala Phe Leu Gly Val Leu Tyr Ala  
 429

545                                      550                                      555                                      560  
 Gly His Gly Arg Leu Ala Ala Ala Thr His Thr Ala Arg Leu Thr Gly  
    565                                      570                                      575  
 Val Thr Ser Leu Val Leu Thr Val Gly Asp Val Asp Arg Met Ser Ala  
 5                                      580                                      585                                      590  
 Phe Asp Arg Gly Pro Ala Gly Ala Ala Gly Arg Thr Arg Thr Ala Gly  
    595                                      600                                      605  
 Tyr Leu Asp Ala Leu Leu Thr Val Cys Leu Ala Arg Ala Gln His Gly  
    610                                      615                                      620  
 10 Gln Ser Val  
 625

## (2) INFORMATION FOR SEQ ID NO:150:

- 15 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 908 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

25 Val Ser Ile Ser Ala Gly Val Arg Gly Gln Gly Trp His Arg Ile Ser  
     1                                      5                                      10                                      15  
 Thr Pro Pro Lys Asn Gly Ala Gly Arg Ser Val Leu Val Phe Gly Leu  
    20                                      25                                      30  
 Val Leu Pro Leu Cys Phe Tyr Pro His Pro Thr Pro Ser Phe Gly Pro  
 30                                      35                                      40                                      45  
 Arg Leu Arg Gln Gln Arg Ala Ser Asp Ser Leu Arg Gly Ala Glu Pro  
    50                                      55                                      60  
 Leu Trp Ala Val Gly Thr Asp Thr Pro Pro Ser Ala Asp Trp Gln Pro  
    65                                      70                                      75                                      80  
 35 Gly Arg Thr Thr Met Gly Pro Gly Leu Trp Val Val Met Gly Val Leu  
    85                                      90                                      95  
 Val Gly Val Ala Gly Gly His Asp Thr Tyr Trp Thr Glu Gln Ile Asp  
    100                                      105                                      110  
 Pro Trp Phe Leu His Gly Leu Gly Leu Ala Arg Thr Tyr Trp Arg Asp  
 40                                      115                                      120                                      125  
 Thr Asn Thr Gly Arg Leu Trp Leu Pro Asn Thr Pro Asp Ala Ser Asp  
    130                                      135                                      140  
 Pro Gln Arg Gly Arg Leu Ala Pro Pro Gly Glu Leu Asn Leu Thr Thr

	145		150		155		160									
	Ala	Ser	Val	Pro	Met	Leu	Arg	Trp	Tyr	Ala	Glu	Arg	Phe	Cys	Phe	Val
				165				170						175		
	Leu	Val	Thr	Thr	Ala	Glu	Phe	Pro	Arg	Asp	Pro	Gly	Gln	Leu	Leu	Tyr
5			180					185					190			
	Ile	Pro	Lys	Thr	Tyr	Leu	Leu	Gly	Arg	Pro	Arg	Asn	Ala	Ser	Leu	Pro
			195					200					205			
	Glu	Leu	Pro	Glu	Ala	Gly	Pro	Thr	Ser	Arg	Pro	Pro	Ala	Glu	Val	Thr
			210				215						220			
10	Gln	Leu	Lys	Gly	Leu	Ser	His	Asn	Pro	Gly	Ala	Ser	Ala	Leu	Leu	Arg
	225				230					235				240		
	Ser	Arg	Ala	Trp	Val	Thr	Phe	Ala	Ala	Ala	Pro	Asp	Arg	Glu	Gly	Leu
			245						250				255			
	Thr	Phe	Pro	Arg	Gly	Asp	Asp	Gly	Ala	Thr	Glu	Arg	His	Pro	Asp	Gly
15			260					265					270			
	Arg	Arg	Asn	Ala	Pro	Pro	Pro	Gly	Pro	Pro	Ala	Gly	Thr	Pro	Arg	His
			275					280					285			
	Pro	Thr	Thr	Asn	Leu	Ser	Ile	Ala	His	Leu	His	Asn	Ala	Ser	Val	Thr
			290					295					300			
20	Trp	Leu	Ala	Arg	Leu	Leu	Arg	Thr	Pro	Gly	Arg	Tyr	Val	Tyr	Leu	Ser
	305				310					315				320		
	Pro	Ser	Ala	Ser	Thr	Trp	Pro	Val	Gly	Val	Trp	Thr	Thr	Gly	Gly	Leu
			325						330				335			
	Ala	Phe	Gly	Cys	Asp	Ala	Ala	Leu	Val	Arg	Ala	Arg	Tyr	Gly	Lys	Gly
25			340					345					350			
	Phe	Met	Gly	Leu	Val	Ile	Ser	Met	Arg	Asp	Ser	Pro	Pro	Ala	Glu	Ile
			355					360					365			
	Ile	Val	Val	Pro	Ala	Asp	Lys	Thr	Leu	Ala	Arg	Val	Gly	Asn	Pro	Thr
			370					375					380			
30	Asp	Glu	Asn	Ala	Pro	Ala	Val	Leu	Pro	Gly	Pro	Pro	Ala	Gly	Pro	Arg
	385				390					395				400		
	Tyr	Arg	Val	Phe	Val	Leu	Gly	Ala	Pro	Thr	Pro	Ala	Asp	Asn	Gly	Ser
			405						410				415			
	Ala	Leu	Asp	Ala	Leu	Arg	Arg	Val	Ala	Gly	Tyr	Pro	Glu	Glu	Ser	Thr
35			420					425					430			
	Asn	Tyr	Ala	Gln	Tyr	Met	Ser	Arg	Ala	Tyr	Ala	Glu	Phe	Leu	Gly	Glu
			435					440					445			
	Asp	Pro	Gly	Ser	Gly	Thr	Asp	Ala	Arg	Pro	Ser	Leu	Phe	Trp	Arg	Leu
			450					455					460			
40	Ala	Gly	Leu	Leu	Ala	Ser	Ser	Gly	Phe	Ala	Phe	Val	Asn	Ala	Ala	His
	465				470					475			480			
	Ala	His	Asp	Ala	Ile	Arg	Leu	Ser	Asp	Leu	Leu	Gly	Phe	Leu	Ala	His
			485						490				495			

Ser Arg Val Leu Ala Gly Leu Ala Arg Ala Ala Gly Cys Ala Ala Asp  
 500 505 510  
 Ser Val Phe Leu Asn Val Ser Val Leu Asp Pro Ala Ala Arg Leu Arg  
 515 520 525  
 5 Leu Glu Ala Arg Leu Gly His Leu Val Ala Ala Ile Arg Glu Gln Ser  
 530 535 540  
 Leu Ala Ala His Ala Leu Gly Tyr Gln Leu Ala Phe Val Leu Asp Ser  
 545 550 555 560  
 Pro Ala Ala Tyr Gly Ala Val Ala Pro Ser Ala Ala Arg Leu Ile Asp  
 10 565 570 575  
 Ala Leu Tyr Ala Glu Phe Leu Gly Gly Arg Ala Leu Thr Ala Pro Met  
 580 585 590  
 Val Arg Arg Ala Leu Phe Tyr Ala Thr Ala Val Leu Arg Ala Pro Phe  
 595 600 605  
 15 Leu Ala Gly Ala Pro Ser Ala Glu Gln Arg Glu Arg Ala Arg Arg Gly  
 610 615 620  
 Leu Leu Ile Thr Thr Ala Leu Cys Thr Ser Asp Val Ala Ala Ala Thr  
 625 630 635 640  
 His Ala Asp Leu Arg Ala Ala Arg Thr Asp His Gln Lys Asn Leu Phe  
 20 645 650 655  
 Trp Leu Pro Asp His Phe Ser Pro Cys Ala Ala Ser Leu Arg Phe Asp  
 660 665 670  
 Leu Ala Glu Gly Gly Phe Ile Leu Asp Ala Met Ala Thr Arg Ser Asp  
 675 680 685  
 25 Ile Pro Ala Asp Val Met Ala Gln Gln Thr Arg Gly Val Ala Ser Val  
 690 695 700  
 Leu Thr Arg Trp Ala His Tyr Asn Ala Leu Ile Arg Ala Phe Val Pro  
 705 710 715 720  
 Glu Ala Thr His Gln Cys Ser Gly Pro Ser His Asn Ala Glu Pro Arg  
 30 725 730 735  
 Ile Leu Val Pro Ile Thr His Asn Ala Ser Tyr Val Val Thr His Thr  
 740 745 750  
 Pro Leu Pro Arg Gly Ile Gly Tyr Lys Leu Thr Gly Val Asp Val Arg  
 755 760 765  
 35 Arg Pro Leu Phe Ile Thr Tyr Leu Thr Ala Thr Cys Glu Gly His Ala  
 770 775 780  
 Arg Glu Ile Glu Pro Lys Arg Leu Val Arg Thr Glu Asn Arg Arg Asp  
 785 790 795 800  
 Leu Gly Leu Val Gly Ala Val Phe Leu Arg Tyr Thr Pro Ala Gly Glu  
 40 805 810 815  
 Val Met Ser Val Leu Leu Val Asp Thr Asp Ala Thr Gln Gln Gln Leu  
 820 825 830  
 Ala Gln Gly Pro Val Ala Gly Thr Pro Asn Val Phe Ser Ser Asp Val  
 432

[illegible]

(2) INFORMATION FOR SEQ ID NO:151:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 370 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

	Met	Ala	Ser	His	Ala	Gly	Gln	Gln	His	Ala	Pro	Ala	Phe	Gly	Gln	Ala
	1				5					10					15	
25	Ala	Arg	Ala	Ser	Gly	Pro	Thr	Asp	Gly	Arg	Ala	Ala	Ser	Arg	Pro	Ser
				20					25					30		
	His	Arg	Gln	Gly	Ala	Ser	Asp	Pro	Glu	Leu	Pro	Thr	Leu	Leu	Arg	Val
			35					40					45			
	Tyr	Ile	Asp	Gly	Pro	His	Gly	Val	Gly	Lys	Thr	Thr	Thr	Ser	Ala	Gln
30		50					55					60				
	Leu	Met	Glu	Ala	Leu	Gly	Pro	Arg	Asp	Asn	Ile	Val	Tyr	Val	Pro	Glu
	65					70					75				80	
	Pro	Met	Thr	Tyr	Trp	Gln	Val	Leu	Gly	Ala	Ser	Glu	Thr	Leu	Thr	Asn
				85						90					95	
35	Ile	Tyr	Asn	Thr	Gln	His	Arg	Leu	Asp	Arg	Gly	Glu	Ile	Ser	Ala	Gly
				100					105					110		
	Glu	Ala	Ala	Val	Val	Met	Thr	Ser	Ala	Gln	Ile	Thr	Met	Ser	Thr	Pro
			115					120					125			
	Tyr	Ala	Ala	Thr	Asp	Ala	Val	Leu	Ala	Pro	His	Ile	Gly	Gly	Glu	Ala
40		130						135				140				
	Val	Gly	Pro	Gln	Ala	Pro	Pro	Pro	Ala	Leu	Thr	Leu	Val	Phe	Asp	Arg
	145					150					155				160	
	His	Pro	Ile	Ala	Ser	Leu	Leu	Cys	Tyr	Pro	Ala	Ala	Arg	Tyr	Leu	Met



165 170 175  
 Gly Ser Met Thr Pro Gln Ala Val Leu Ala Phe Val Met Pro Pro Thr  
 180 185 190  
 Ala Pro Gly Thr Asn Leu Val Leu Gly Val Leu Pro Glu Ala Glu His  
 5 195 200 205  
 Ala Asp Arg Leu Ala Arg Arg Gln Arg Pro Gly Glu Arg Leu Asp Leu  
 210 215 220  
 Ala Met Leu Ser Ala Ile Arg Arg Val Tyr Asp Leu Leu Ala Asn Thr  
 225 230 235 240  
 10 Val Arg Tyr Leu Gln Arg Gly Gly Arg Trp Arg Glu Asp Trp Gly Arg  
 245 250 255  
 Leu Thr Gly Val Ala Ala Ala Thr Pro Arg Pro Asp Pro Glu Asp Gly  
 260 265 270  
 Ala Gly Ser Leu Pro Arg Ile Glu Asp Thr Leu Phe Ala Leu Phe Arg  
 15 275 280 285  
 Val Pro Glu Leu Leu Ala Pro Asn Gly Asp Leu Tyr His Ile Phe Ala  
 290 295 300  
 Trp Val Leu Asp Val Leu Ala Asp Arg Leu Leu Pro Met His Leu Phe  
 305 310 315 320  
 20 Val Leu Asp Tyr Asp Gln Ser Pro Val Gly Cys Arg Asp Ala Leu Leu  
 325 330 335  
 Arg Leu Thr Ala Gly Met Ile Pro Thr Arg Val Thr Thr Ala Gly Ser  
 340 345 350  
 Ile Ala Glu Ile Arg Asp Leu Ala Arg Thr Phe Ala Arg Glu Val Gly  
 25 355 360 365  
 Gly Val  
 370

## (2) INFORMATION FOR SEQ ID NO:152:

30

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 352 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

35

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

40

Val Leu Arg Val Val Asp Val Arg Gln Gly Leu Gly Gly Pro Gln His  
 1 5 10 15  
 Leu Pro Val Ser His Arg Leu Gly Asp Val Asp Asp Ile Val Ala Arg  
 434

	20	25	30
	Pro Gln Gly Leu His Gln Leu Arg Gly Gly Gly Gly Leu Pro His Pro		
	35	40	45
	Val Gly Ser Val Tyr Ile Asn Pro Gln Gln Arg Gly Gln Leu Arg Ile		
5	50	55	60
	Pro Ala Gly Phe Gly Gly Pro Leu Ala Met Ala Arg Thr Gly Arg Arg		
	65	70	75
	Ala Ala Val Gly Arg Pro Ala Arg Thr Ser Ser Leu Thr Glu Arg Arg		
	85	90	95
10	Arg Val Leu Leu Ala Gly Val Arg Ser His Thr Arg Phe Tyr Lys Ala		
	100	105	110
	Phe Ala Arg Glu Val Arg Glu Phe Asn Ala Thr Arg Ile Cys Gly Thr		
	115	120	125
	Leu Leu Thr Leu Met Ser Gly Ser Leu Gln Gly Arg Ser Leu Phe Glu		
15	130	135	140
	Ala Thr Arg Val Thr Leu Ile Cys Glu Val Asp Leu Gly Pro Arg Arg		
	145	150	155
	Pro Asp Cys Ile Cys Val Phe Glu Phe Ala Asn Asp Lys Thr Leu Gly		
	165	170	175
20	Gly Val Cys Val Ile Leu Lys Thr Cys Lys Ser Ile Ser Ser Gly Asp		
	180	185	190
	Thr Ala Ser Lys Arg Glu Gln Arg Thr Thr Gly Met Lys Gln Leu Arg		
	195	200	205
	His Ser Leu Lys Leu Leu Gln Ser Leu Ala Pro Pro Gly Asp Lys Val		
25	210	215	220
	Val Tyr Leu Cys Pro Ile Leu Val Phe Val Ala Gln Arg Thr Leu Arg		
	225	230	235
	Val Ser Arg Val Thr Arg Leu Val Pro Gln Lys Ile Ser Gly Asn Ile		
	245	250	255
30	Thr Ala Ala Val Arg Met Leu Gln Ser Leu Ser Thr Tyr Ala Val Pro		
	260	265	270
	Pro Glu Pro Gln Thr Arg Arg Ser Arg Arg Arg Val Ala Ala Thr Ala		
	275	280	285
	Arg Pro Gln Arg Pro Pro Ser Pro Thr Arg Asp Pro Glu Gly Thr Ala		
35	290	295	300
	Gly His Pro Ala Pro Pro Glu Ser Asp Pro Pro Ser Pro Gly Val Val		
	305	310	315
	Gly Val Ala Ala Glu Gly Gly Gly Val Leu Gln Lys Ile Ala Ala Leu		
	325	330	335
40	Phe Cys Val Pro Val Ala Ala Lys Ser Arg Pro Arg Thr Lys Thr Glu		
	340	345	350

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 571 amino acids

(B) TYPE: amino acid

5 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

```

Met Asp Pro Tyr Tyr Pro Phe Asp Ala Leu Asp Val Trp Glu His Arg
 1           5           10           15
Arg Phe Ile Val Ala Asp Ser Arg Ser Phe Ile Thr Pro Glu Phe Pro
15           20           25           30
Arg Asp Phe Trp Met Leu Pro Val Phe Asn Ile Pro Arg Glu Thr Ala
           35           40           45
Ala Glu Arg Ala Ala Val Leu Gln Ala Gln Arg Thr Ala Ala Ala Ala
           50           55           60
20 Ala Leu Glu Asn Ala Ala Leu Gln Ala Ala Glu Leu Pro Val Asp Ile
           65           70           75           80
Glu Arg Arg Ile Arg Pro Ile Glu Gln Gln Val His His Ile Ala Asp
           85           90           95
Ala Leu Glu Ala Leu Glu Thr Ala Ala Ala Ala Glu Glu Ala Asp
25           100          105          110
Ala Ala Arg Asp Ala Glu Arg Glu Gly Ala Ala Asp Gly Ala Ala Pro
           115          120          125
Ser Pro Thr Ala Gly Pro Ala Ala Ala Glu Met Glu Val Gln Ile Val
           130          135          140
30 Arg Asn Asp Pro Pro Leu Arg Tyr Asp Thr Asn Leu Pro Val Asp Leu
           145          150          155          160
Leu His Met Val Tyr Ala Gly Arg Gly Ala Ala Gly Ser Ser Gly Val
           165          170          175
Val Phe Gly Thr Trp Tyr Arg Thr Ile Gln Glu Arg Thr Ile Ala Asp
35           180          185          190
Phe Pro Leu Thr Thr Arg Ser Ala Asp Phe Arg Asp Gly Arg Met Ser
           195          200          205
Lys Thr Phe Met Thr Ala Leu Val Leu Ser Leu Gln Ser Cys Gly Arg
           210          215          220
40 Leu Tyr Val Gly Gln Arg His Tyr Ser Ala Phe Glu Cys Ala Val Leu
           225          230          235          240
Cys Leu Tyr Leu Leu Tyr Arg Thr Thr His Glu Ser Ser Pro Asp Arg
           245          250          255

```

Asp Arg Ala Pro Val Ala Phe Gly Asp Leu Leu Ala Arg Leu Pro Arg  
 260 265 270  
 Tyr Leu Ala Arg Leu Ala Ala Val Ile Gly Asp Glu Ser Gly Arg Pro  
 275 280 285  
 5 Gln Tyr Arg Tyr Arg Asp Asp Lys Leu Pro Lys Ala Gln Phe Ala Ala  
 290 295 300  
 Ala Gly Gly Arg Tyr Glu His Gly Ala Thr His Val Val Ile Ala Thr  
 305 310 315 320  
 Leu Val Arg His Gly Val Leu Pro Ala Ala Pro Gly Asp Val Pro Arg  
 10 325 330 335  
 Asp Thr Ser Thr Arg Val Asn Pro Asp Asp Val Ala His Arg Asp Asp  
 340 345 350  
 Val Asn Arg Ala Ala Ala Ala Phe Leu Arg His Asn Leu Phe Leu Trp  
 355 360 365  
 15 Glu Asp Gln Thr Leu Leu Arg Ala Thr Ala Asn Thr Ile Thr Ala Val  
 370 375 380  
 Leu Arg Arg Leu Leu Ala Asn Gly Asn Val Tyr Ala Asp Arg Leu Asp  
 385 390 395 400  
 Asn Arg Leu Gln Leu Gly Met Leu Ile Pro Gly Ala Val Pro Ala Glu  
 20 405 410 415  
 Ala Ile Arg Ala Ser Gly Leu Asp Ser Gly Ala Ile Lys Ser Gly Asp  
 420 425 430  
 Asn Asn Leu Glu Ala Leu Cys Val Asn Tyr Val Leu Pro Leu Tyr Gln  
 435 440 445  
 25 Ala Asp Pro Thr Val Glu Leu Thr Gln Leu Phe Pro Gly Leu Ala Ala  
 450 455 460  
 Leu Cys Leu Asp Ala Gln Ala Gly Arg Pro Leu Ala Ser Thr Arg Arg  
 465 470 475 480  
 Val Val Asp Met Ser Ser Gly Ala Arg Gln Ala Ala Leu Val Arg Leu  
 30 485 490 495  
 Thr Ala Leu Glu Leu Ile Asn Arg Thr Arg Thr Asn Thr Thr Pro Val  
 500 505 510  
 Gly Glu Ile Ile Asn Ala His Asp Ala Leu Gly Ile Gln Tyr Glu Gln  
 515 520 525  
 35 Gly Leu Gly Leu Leu Ala Gln Gln Ala Arg Ile Gln Ala Lys Arg Phe  
 530 535 540  
 Ala Thr Phe Asn Val Gly Ser Asp Tyr Asp Leu Leu Tyr Phe Leu Cys  
 545 550 555 560  
 Leu Gly Phe Ile Pro Gln Tyr Leu Ser Val Ala  
 40 565 570

(2) INFORMATION FOR SEQ ID NO:154:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 571 amino acids.

(B) TYPE: amino acid.

(C) STRANDEDNESS: single

5

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

10

```

Met Asp Pro Tyr Tyr Pro Phe Asp Ala Leu Asp Val Trp Glu His Arg
 1           5           10           15
Arg Phe Ile Val Ala Asp Ser Arg Ser Phe Ile Thr Pro Glu Phe Pro
          20           25           30
15 Arg Asp Phe Trp Met Leu Pro Val Phe Asn Ile Pro Arg Glu Thr Ala
          35           40           45
Ala Glu Arg Ala Ala Val Leu Gln Ala Gln Arg Thr Ala Ala Ala Ala
          50           55           60
Ala Leu Glu Asn Ala Ala Leu Gln Ala Ala Glu Leu Pro Val Asp Ile
20 65           70           75           80
Glu Arg Arg Ile Arg Pro Ile Glu Gln Gln Val His His Ile Ala Asp
          85           90           95
Ala Leu Glu Ala Leu Glu Thr Ala Ala Ala Ala Glu Glu Ala Asp
          100          105          110
25 Ala Ala Arg Asp Ala Glu Arg Glu Gly Ala Ala Asp Gly Ala Ala Pro
          115          120          125
Ser Pro Thr Ala Gly Pro Ala Ala Ala Glu Met Glu Val Gln Ile Val
          130          135          140
Arg Asn Asp Pro Pro Leu Arg Tyr Asp Thr Asn Leu Pro Val Asp Leu
30 145          150          155          160
Leu His Met Val Tyr Ala Gly Arg Gly Ala Ala Gly Ser Ser Gly Val
          165          170          175
Val Phe Gly Thr Trp Tyr Arg Thr Ile Gln Glu Arg Thr Ile Ala Asp
          180          185          190
35 Phe Pro Leu Thr Thr Arg Ser Ala Asp Phe Arg Asp Gly Arg Met Ser
          195          200          205
Lys Thr Phe Met Thr Ala Leu Val Leu Ser Leu Gln Ser Cys Gly Arg
          210          215          220
Leu Tyr Val Gly Gln Arg His Tyr Ser Ala Phe Glu Cys Ala Val Leu
40 225          230          235          240
Cys Leu Tyr Leu Leu Tyr Arg Thr Thr His Glu Ser Ser Pro Asp Arg
          245          250          255
Asp Arg Ala Pro Val Ala Phe Gly Asp Leu Leu Ala Arg Leu Pro Arg

```

		260		265		270	
		Tyr Leu Ala Arg Leu Ala Ala Val Ile Gly Asp Glu Ser Gly Arg Pro					
		275		280		285	
		Gln Tyr Arg Tyr Arg Asp Asp Lys Leu Pro Lys Ala Gln Phe Ala Ala					
5		290		295		300	
		Ala Gly Gly Arg Tyr Glu His Gly Ala Thr His Val Val Ile Ala Thr					
		305		310		315	320
		Leu Val Arg His Gly Val Leu Pro Ala Ala Pro Gly Asp Val Pro Arg					
			325		330		335
10		Asp Thr Ser Thr Arg Val Asn Pro Asp Asp Val Ala His Arg Asp Asp					
			340		345		350
		Val Asn Arg Ala Ala Ala Ala Phe Leu Arg His Asn Leu Phe Leu Trp					
			355		360		365
		Glu Asp Gln Thr Leu Leu Arg Ala Thr Ala Asn Thr Ile Thr Ala Val					
15		370		375		380	
		Leu Arg Arg Leu Leu Ala Asn Gly Asn Val Tyr Ala Asp Arg Leu Asp					
		385		390		395	400
		Asn Arg Leu Gln Leu Gly Met Leu Ile Pro Gly Ala Val Pro Ala Glu					
			405		410		415
20		Ala Ile Arg Ala Ser Gly Leu Asp Ser Gly Ala Ile Lys Ser Gly Asp					
			420		425		430
		Asn Asn Leu Glu Ala Leu Cys Val Asn Tyr Val Leu Pro Leu Tyr Gln					
			435		440		445
		Ala Asp Pro Thr Val Glu Leu Thr Gln Leu Phe Pro Gly Leu Ala Ala					
25		450		455		460	
		Leu Cys Leu Asp Ala Gln Ala Gly Arg Pro Leu Ala Ser Thr Arg Arg					
		465		470		475	480
		Val Val Asp Met Ser Ser Gly Ala Arg Gln Ala Ala Leu Val Arg Leu					
			485		490		495
30		Thr Ala Leu Glu Leu Ile Asn Arg Thr Arg Thr Asn Thr Thr Pro Val					
			500		505		510
		Gly Glu Ile Ile Asn Ala His Asp Ala Leu Gly Ile Gln Tyr Glu Gln					
			515		520		525
		Gly Leu Gly Leu Leu Ala Gln Gln Ala Arg Ile Gln Ala Lys Arg Phe					
35		530		535		540	
		Ala Thr Phe Asn Val Gly Ser Asp Tyr Asp Leu Leu Tyr Phe Leu Cys					
		545		550		555	560
		Leu Gly Phe Ile Pro Gln Tyr Leu Ser Val Ala					
			565		570		
40							

(2) INFORMATION FOR SEQ ID NO:155:

(i) SEQUENCE CHARACTERISTICS:

439

(A) LENGTH: 11706 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

	GGACGAACCA	ACGGACACCT	CCGCAAAGCG	CGCGCGCGCC	TCCCCGCGG	CGTCGCGACA	60
10	GACCAGATAC	AGCAGGGCGT	GGAGGCAGTC	GCGCGTGCGC	GGGGGCAGCC	ATACCGCGTA	120
	TAGGGTAATG	GCGCTGACGC	TCTCCTCCAC	CCAAACGATG	CCGGGGGCTT	CCATGCCACG	180
	ACGCCCCGGG	GTTGCCGTGT	ATCGAACGAG	CGCGGCCCCA	GACTTATAGG	GTGCTAAAGT	240
	TCACCGCCCC	CTGCATCATG	GGCCAGGCCT	CGGTGGGAAG	CTCCGACAGA	GCCGCCTCGA	300
	GAATGATGTC	AGTGTGGGC	TGGGCGCCGG	AGGCGTGCGT	GCGCAAGCAG	CGCCCCACG	360
15	CGGGCGCGCG	CAGCTTGAAG	CGCGCGCCCG	CAAACCTCCG	CTTATGGGCC	ATCAGCAGCG	420
	CGTACAGCTG	TCTGTGCGTC	CGGCAGGCGC	TGTGGTTCGAT	GCGGTGGGCG	TCCAGCAGCT	480
	CCACGATGGC	TCGCTTGGTG	AGGTTTTTAA	CGCGCCCCGC	CCCGGGAAC	GTCTGCGTGC	540
	TCTTGGCCAG	CTGCACCCCG	AACAGTTCGC	CCCAGATGAT	CTTGAACAGC	GACAGCGCGT	600
	GCTCCGTCTC	GCTCACGGAC	CCGCGCGGGG	GGCAGCCGCT	CAGGGCGTCG	GCCACGCGCT	660
20	TAACCGCGTC	CTCCGACAGC	AAGGGGCCGT	CGGTCACGTT	ACAGTGGCCC	AGTTCGAACA	720
	CCAGCTGCAT	GTAGCGGTCG	TAGTGGGGGT	TCAGCAGCTC	CAGCACGTCC	TCGGGGCTAA	780
	AGGTTCGCCC	CGACCCCCCG	GCCATCGAGT	CCCACTGCAG	GCACGCGGCC	ATGGTGTCTG	840
	ACAGACGGAA	CAGCTCCCAG	ACGGGGGCGA	CGTTTAGGGT	GGGGTGTAGG	GCCACAAGCT	900
	CCAGCTCTCC	GGCGGCGTTG	ATCGTGGGGA	TGACGCCCCG	GGCGTAGTGG	TCGTAAAGCC	960
25	GCCGGAAGAT	GGCGCTGCTA	TGGGCGGCCA	TGGGGACGCG	AAGACAGGCC	TCCAGCAGCA	1020
	CCAGGTAGAT	GAACCGCGTG	CGGCCGACCA	GGCTGTTGAG	GCCGCGCATG	AGCGCGACCA	1080
	CCTCGGCCGG	CGCGACGTCC	GGCCGGAGGT	ACTTTTCGAC	GAAAAGGCCC	ACCTCTCCG	1140
	TCTCGCGGCG	CTGGGCCGAC	AGGGACGTGT	CGGGGTCTTG	GCAGCGCAGC	TCCCGCAGAT	1200
	CCCGCTGGGC	CCTCAGGGCA	TCAAAATGTA	TCCCCGCAA	AAACAGACAA	AAGTTCCTCG	1260
30	GGGTACGCGC	GGCGTCGTGG	CCCCAGAACC	GCACGTGCAT	GCAGTTGAGG	GTCAGAAGCA	1320
	TGTGGAGGAT	GTTAAGACTG	TCCGCGAGGC	ACGCCAGCGT	GCACCTCTCG	AAGTAGTGCT	1380
	TGTACCGGAA	TTTGCTGTAG	ATGCGCGACC	CCCGCGCCTG	CGCCGCGTCG	GCGTGCGACG	1440
	CGTCGACGCG	CCCTTTGAAC	CGGCGGCACA	ACAGGTTTCGT	CACCTGGGAA	AACTGTGCCG	1500
	GCCACTGCCC	GCTGGCGCTC	ACCACGTGGT	TGAGCAGCAT	GGGCGTAAAG	ACGGGCTCCG	1560
35	AGCGCGCCCC	GGACCCGTCC	ATGTAGATCA	GCAGCTCCCC	CTTGCGGAGA	GTCCGTACCC	1620
	GCCCCAGCGA	CTGGTACACG	GACACCATGT	CCGGCCCCGA	GTTTATGGGT	TTCACGTAGG	1680
	CGAACATGCT	GTCAAAGTGC	GGCGGATCGA	AGCTAAGGCC	CACCGTCACG	ACCGTTGTGT	1740
	AGATGACCAC	CCGGTACCGG	CCCCATGTGG	TCACGTCCGC	GGGCGGGGTG	AGCGAGTGGA	1800
	GCAGCAGCAC	GCGGTCCGTA	AACTGCCGGC	AGAACCTGGC	AACGACCTCC	GCGAAGGAGA	1860
40	CCGTCGACGA	GAAGATGCAG	ACGTTATCTC	CGCCGGCCAG	GCGCGCCTCC	AGCTCCCCGA	1920
	AGAAGGTGGC	GTCCGGGGGG	GCGTCCGGGG	GGGGCGCCCC	GCCCGCCGGC	CCCGGCGGGC	1980
	GCAGGGCCGC	CTGCAGGACC	TCGGGCCCCA	GGCGCGGGAG	AAACAGACAA	CGGCGCGCCG	2040
	AAAATCCGGG	CATGGCATACT	TCCCCGATGA	CCACGTGAAC	GTTCTTTTCG	CCCCGAGGGC	2100

	TGCACAGAAA	GTCCACCAGC	TGCGCGTTGG	CGGTGGCGTC	CATGGCGATG	ATCCGCGGGC	2160
	ACGTGCGCAG	CAGGCGCAGC	ATCAACGCGT	CGACGCGGCC	CAGCTGCTGC	ATCGTCGGCG	2220
	AGTACAGTTG	GCCCAACGTC	GACATGACTT	CGTCCAGGAC	GAGCACGTCG	TAGTTGTTCA	2280
	ACAGGTTCCG	GCCACGCGA	TGAAGACTTT	CCACCTGCAC	GATGAGACGG	TGGAAGGGGC	2340
5	GGTCGTTTCAT	GATGTAATTG	GTGGATGAGA	AGTAGGTGAC	GAAGTCGGGC	AACCTGACT	2400
	CAGCGAACCG	CGTCGCCAGG	GTCTGAGTAA	AACTCCGACG	ACAGGAGACG	ACCAGCACAC	2460
	TCGTGTCCGG	AGAGTGGATC	GCTTCCCCCA	ACCAGCGGAT	CAGCGCGGTA	GTTTTTCCCG	2520
	AGCCCATTTG	CGCGCGGACC	ACAGTTACGC	ACCGGGCCGT	CGGGGCGCTC	GCGTCCGGGA	2580
	AGGTGACGGG	TCCGTGTTGC	TGCCGCTCGA	TCGTTGTTTT	CGGGTGGACC	CGGGGAACCC	2640
10	ACTCGGCCAA	ATCCCCCCCG	TAAAGCATCC	GCGCCAGCGA	TACACTCGAC	GTGTACTGCT	2700
	CGCACTCGTC	ATCCCCGATG	GGACGCCGGG	CCCCCAGGGG	ATCCCCGAG	GCCGCCCCGG	2760
	GCGCCGACGT	CGCGCCCCGG	GCGCGGGCGG	CGTGGTGGGT	CTGGTGTGTG	CAGGTGGCGA	2820
	CGTTCATCGT	CTCGGCCATC	TGCGTCGTGG	GGCTCCTGGT	GCTGGCCTCT	GTGTTCCGGG	2880
	ACAGGTTTCC	CTGCCTTTAC	GCCCCCGCGA	CCTCTTATGC	GGAGGCGAAC	GCCACGGTCG	2940
15	AGGTGCGCGG	GGGTGTAGCC	GTCCCCCTCC	GGTTGGACAC	GCAGAGCCTG	CTGGCCACGT	3000
	ACGCAATTAC	GTCTACGCTG	TTGCTGGCGG	CGGCCGTGTA	CGCCGCGGTG	GGCGCGGTGA	3060
	CCTCGCGCTA	CGAGCGCGCG	CTGGATGCGG	CCCGTCGCCT	GGCGGCGGCC	CGTATGGCGA	3120
	TGCCACACGC	CACGCTAATC	GCCGGAAACG	TCTGCGCGTG	GCTGTTGCAG	ATCACAGTCC	3180
	TGCTGCTGGC	CCACCGCATC	AGCCAGCTGG	CCCACCTTAT	CTACGTCTTG	CAC'TTTCGT	3240
20	GCCTCGTGTA	TCTCGCGGCC	CATTTTTCGA	CCAGGGGGGT	CCTGAGCGGG	ACGTACCTGC	3300
	GTCAGGTTCA	CGGCCTGATT	GACCCGGCGC	CGACGCACCA	TCGTATCGTC	GGTCCGGTGC	3360
	GGGCAGTAAT	GACAAACGCC	TTATTACTGG	GCACCCTCCT	GTGCACGGCC	GCCGCCGCGG	3420
	TCTCGTTGAA	CACGATCGCC	GCCCTCAACT	TCAACTTTTC	CGCCCCGAGC	ATGCTCATCT	3480
	GCCTGACGAC	GCTGTTTCGCC	CTGCTTGTGC	TGTCGCTGTT	GTTGGTGGTC	GAGGGGGTGC	3540
25	TGTGTCACTA	CGTGCGCGTG	TTGGTGGGCC	CCCACCTCGG	GGCCATCGCC	GCCACCGGCA	3600
	TCGTGCGCCT	GGCCTGCGAG	CACTACCACA	CCGGTGGCTA	CTACGTGGTG	GAGCAGCAGT	3660
	GGCCGGGGGC	CCAGACGGGA	GTCCGCGTCG	CCCTGGCGCT	CGTCGCCGCC	TTTGCCCTCG	3720
	CCATGGCCGT	GCTTCGGTGC	ACGCGCGCCT	ACCTGTATCA	CCGGCGACAC	CACACTAAAT	3780
	TTTTCGTGCG	CATGCGCGAC	ACCCGGCACC	GCGCCCATTC	GGCGCTTCGA	CGCGTACGCA	3840
30	GCTCCATGCG	CGGTTCTAGG	CGTGCGGGGC	CGCCCGGAGA	CCCGGGCTAC	GCGGAAACCC	3900
	CCTACGCGAG	CGTGTCACAC	CACGCCGAGA	TCGACCGGTA	TGGGGATTCC	GACGGGGACC	3960
	CGATCTACGA	CGAAGTGGCC	CCCGACCACG	AGGCCGAGCT	CTACGCCCGA	GTGCAACGCC	4020
	CCGGGGCCTGT	GCCCGACGCC	GAGCCCATTT	ACGACACCGT	GGAGGGGTAT	GCGCCAAGGT	4080
	CCGCGGGGGA	GCCGGTGTAC	AGCACCGTTC	GGCGATGGTA	GCCGTTTCGT	TCGTTTTAAT	4140
35	AAACCGACGT	TGTGCGTTTC	ACCATACTTC	GGCGCGCGCG	TGTGTGTGTT	TTTTTTGTGG	4200
	TGTTTTATTTT	CCCCCACCCC	TTCTTTTCT	TTGCGCCACC	ACCCCCCTCC	TCCCCCGTAC	4260
	TATACAACAA	AAAATACCAC	ACATACGACC	AAATACGGAC	AATCATTTCT	GTCTTTATTC	4320
	GCTGTACAGAG	AGTGGGGGCG	TGAGCGTGGC	AGGAGGGCGG	GCCACGTCGG	GGTCCCGCCG	4380
	TCTGGTGTGA	CGCGATGGGG	GGTCCGATGC	GCGCCGGTAC	TGGGGCCCCG	GCGCCCGGGT	4440
40	GACCACGCGC	ATGTCGGGGG	GCACGTAGAA	GTTACCCTCT	TCTTCGGAAT	CGATGTCCAC	4500
	GACGTCAAAT	TCGTGGGCGG	TCAGCGAGAC	GACCTCCCCG	CCGTCCGTGA	TGATGACGTT	4560
	GTGTGCGCAG	CAGCAGGGCC	GCGCCCCGGA	GAACGCGAGG	CCCATAACTT	GGCGAGCGTA	4620
	TCGTGCAAGG	CCAGGCGGCT	GTTTCGCCCG	ATGTCCCGGT	AGATCCCCGG	CTCGACGCGG	4680



	ACGGGGGTGA	TGATCAGGGC	GATCGGAACG	GCCTGGTCCG	GGAGGATCGA	TGCCTTGGCG	4740
	GGTCCGGGGG	CCCCGCCACG	CCCGGCGGGC	GCTCCGCGGC	CGTCTCCAG	GCGGAACGTC	4800
	ACGCCCTCCT	CCGCGCCCCG	GCGGTGCCTG	CCGAGGAACG	TCACCAGGTG	CGGTTGCAGG	4860
	GGGAGTCGG	GAAAGTGGCT	GTCGAGGACG	TATCCCTGCA	CCAAGATCTG	TTTGAAGTTC	4920
5	GGGTGGCGGG	GGTTGGCGAA	GATGGGCTCG	CGGCGAACCA	GCTCCCCGGA	GCTCCAGGCC	4980
	ACGGGAGAGA	TGGTGCGACG	CTCGAGGTG	GGGACGCCAA	ACAGAAGCAC	CTCCGAGACA	5040
	ACGCCGCTAT	TTAACTCCAC	CAGCGCCCGA	TCCGGGGCGG	AGCATCGCCT	TTTTTCGCCG	5100
	GCGGCGCGGG	AATCGAGCCA	GTCCCGGTCT	TGGGTGACGA	GCGCCTCCTC	CGGGCCCCGG	5160
	ACGCGCCCCG	GCGCGAAGTA	GCGCACGCCG	GGGTTGGGGA	TGGACCGGAT	GAACGCCCCG	5220
10	AACGCCTCCG	GCGATCGCCG	CGCCATCAGG	TCCTCGTACG	CGGAGGCCGC	GGGGGCGCCG	5280
	GGGTCCGCGG	GGTCGAACGC	GTACTTGGCT	CGGCACTTAA	CCTCGTAGAA	GGCCAGGGGG	5340
	GTCTGGGGGG	CGGGGGCCAG	GTAGCCGTGA	GGGTCCCTGG	GGCACACGAG	GATGTCCAGG	5400
	GACGCCCCCA	CCATGCCCGT	GTGGCCGTCC	ATGAGGACCC	CGCACGCGTG	CACGTTCTCC	5460
	TCGCGGAGGT	CCCCGGGTG	GTGAAAGACG	AAGCGCCCGG	CGTCGGCGTC	GTCGTTGACG	5520
15	CCCGCGTCCG	CGCGGCCAC	GCAGTAGCGA	AACAGCAGGT	TTGGGGCCGT	CGGCTCGTTC	5580
	ACCCGCCCCG	ACATCACCGC	CGACGACTGG	GCGTCCAGCC	GCAGGCTGGC	GTTGTGGGTG	5640
	AGCCACTGGG	ACGAGAAGCA	CGGACCCTGC	GCGCCCCACC	GCAGCGTGGA	GGCGGTGTC	5700
	AGGCCCCGCC	GAAGCAGGGC	CCAGAGCTGG	CAGTCGGCCT	GGTTTTGCGT	CGCCGCCTCG	5760
	TAAAATCCCA	TAAGCGGGCG	GGGGGCGACG	GCTTCGGCGG	CGGACGGGGG	GGCGCGGCGC	5820
20	GTCAGGCGCC	AGAGGTGCCG	GCCGAGCCCG	CGGTCCACCA	TGCCGGCCGC	CTCCAGCGAC	5880
	ACGACGAGGG	AGCACAGATA	GTCCAGGCCA	GCCCACAGGG	GCCCAGTGCC	CAGAGGGGAG	5940
	CGGACGCCGC	GCAGCAGGCC	GCGCAGGTGG	CGCTCGAACG	TTTCCGCCAA	GATATGGGGG	6000
	GGCAGTGCGT	TGGGGATCGC	CGACGCCGAC	CACATCGGGT	CGGGGTCCGG	GGGACCGGGG	6060
	CTGCAGTCCG	GGTCGATGGC	GTGTGCGCCC	CCCGGCGAGA	GGGGAATGTC	GGGGGTGGGC	6120
25	GGGCCGGATG	AGGCCTCAGA	GAGGGCCGGG	GACGCGGGCC	GGGCCTTTTC	GCCCCGGGCC	6180
	CCGCCGTCCG	GTTGCCACAG	TGGGGGGCTC	TGGGGCCAAT	GGGAACCCGG	GGCCCCCGGT	6240
	GACGTGGGGC	GGGGTGGGGC	GGGGCGGGCC	CAAAGACCGG	TCGCCAGATC	TAGGCTGTTG	6300
	GGTCGGGGCC	GCTTCGGGGG	ACTATCGGGG	TCGCGGGCGG	GGTCCGCGGG	GCGCTTGGCG	6360
	CCGGGTGTTG	CGGCGGCCGC	CATTTTACG	AGCAGCCGAA	GAGCTCGAGG	GCGGAAGGGA	6420
30	TCCTCACGAC	AGAGAGTGGC	GCGCGGCCGG	GTGGCGTGA	CAGAGGCGGG	AGACCAGCAC	6480
	CAGCAGCGGC	CTCAGCTCGG	GCGGCAGCGA	CACCGACGAC	AGGACGGCCT	TGTGCGTGCG	6540
	CTGGTAATTT	ATACACTGCT	CCGTGAACGC	GCGCCGAATC	TTGGGATTGC	GAAGGTGGCG	6600
	CCGGATGCCC	TCCGGCACGT	CATACGCCAG	GCCGTGGGTG	TTGGTCTCGG	CCGAGTTGAC	6660
	AAAGAGGGCG	GGGTGCAGAA	CGCAGCGATA	GGCGAGGAGG	GCCACGGCAA	AGTCCGGCGA	6720
35	GAGCTGGTTG	TTAAAGTACT	GGTAGCCCGG	GACGCGGGTC	ACGGGGACGC	CCAGGCTCGG	6780
	GGCCACGTAC	ACGCTAACCA	GCAGCTCCAG	CAGCGTCTGC	CCCAGGGCGT	AGAGATCGAC	6840
	CGCCAGCCCG	ACGTCGTGCT	TCAGGGGGCG	GTTGTTAAAC	TCGGCCCGCT	CGTTGTTGAG	6900
	GTACTTTACC	AAGAGCTCCG	GCGGCTGGTT	GTACCCGTGC	CCCACCAGAG	TGTGAAAGTT	6960
	GGCCGTGGTC	AGGGCGGCGG	GCATCCCAAA	CCCCCGGGGG	GACTCGAGGT	CCGGCTCCTG	7020
40	GAGGCAAAAC	TGGCCCCGGG	ATATCGTGGA	GTTGGAGTTC	AGGGTCACCA	GGCTAAAGTC	7080
	GGCCAGGACG	GCCCCCGGA	GCGACACCGC	GTCCGATCGC	AGCATCACGA	GGACGTTGGC	7140
	GCACTTGATG	TCCAGGTGGC	TGATCCCGCA	CCTGGTGTTC	AGGAACACCA	CGGCGCGCGC	7200
	CAGGTCTGTG	AAGCAGTGGT	GGAGGGCCGT	GCGACGAGAG	GGGGTGGTCG	CGCGCAGGGA	7260

	CGCCAGCTGG	CCGATGTACT	TGCCGAGGTC	CATGTCGTAC	GCGGGGAACA	CGATCTGGCG	7320
	CTGCTGCAGC	GAGAACCCGA	GCGGGGTGAT	AAAGCCGCGG	ATGTCGTGGG	TGCGGCCGCC	7380
	GCGAAGAGCG	CACTCCCCCA	CGAGCAGGGT	CGCGACGAGC	TCCACGGCAA	ACCACTCCTT	7440
	TTCCCGGATG	GTCTTCACGG	CGAGCTTGTG	TTCGCGAATC	AACTGCACCT	CGCCGTACCC	7500
5	CCCCGAGCCC	CCGAAGCTGC	GGGCCCCGGG	GATCTCCAGG	GTCGTGTAGC	GGAGGGCGGG	7560
	GTTGACGGCG	AATACGGGGA	TGCATAGCTT	GTGGATGCGC	GCGAGGGACA	GGATGTGCGA	7620
	GGGGGGCGAC	GGGGGCGAGG	TCATGGCCGT	CTCGGACCTG	GCGAGGGGCG	GGCGCCTCAG	7680
	CTTGCCCGCA	GGGCCGGGGG	CCTCGGGGGA	CGAGCGGCGA	CGAGACGAGC	GGCTCACTCG	7740
	CCATCGGGAC	AGTCCCGCGC	GAAGCCGCTC	CCGGAAGCTG	GATCGGCGGC	GGGACCCGGG	7800
10	GCGGGCTCCG	GAGACGGCGC	CGTCTCGGGG	GGAGGGGCCG	CTTGGGCGTC	CGGACGCCCG	7860
	GCGGCTGAGG	GAGTGTATGT	AGGACCGGAG	CCAGGCCTTG	AAGGAGCGTC	GGTGTGCACC	7920
	TTGGGGGCTG	ATGTCAGCTG	CCACATGACT	AGCAGGTCGC	TGTCGCCCCG	ACTCATCCAT	7980
	CCGTCCGCCA	GGTCGCCGTC	CCCCACAGA	GACCGGTTTC	CCGCGGCCCTC	TTGAGCTGTC	8040
	TCCTCCTGGT	CCGCAAGACG	ATCGTCCGCC	GCGTCCAGGC	GCTCGCTAAG	CGCGGGATCG	8100
15	AGGTACCGTC	GGTGTGCGGT	TAGAAATCA	CGTCGCGCCG	CTTGCTCTTC	CACGCGAATT	8160
	TTAACACAGG	TCGCTCGCTG	TCGCATCATC	TCTAAGCGCG	CGCGGGACTT	TAGCCGCGCC	8220
	TCCAATTCCA	AGTGGGCCGC	CTTGGCGGCC	ATAAAGCGCG	CAACAAACCT	AGGATCTTGT	8280
	GTACTCACGC	CCTCCCGGTG	TAGCTGCAGG	GTCTGGTCCC	TGTACACCTC	GGCCCGGAGG	8340
	TGCGTCTCGG	CCAAACGTGG	GCGCAGGGCC	GCGTGGCTGG	CGTCTCGGCT	CATCTCGCCG	8400
20	CCCCCGCGCG	CGCCCGACGT	CGGACTCCTT	CGCCCCGACC	CCCCTGACCT	CAGCCGCCCC	8460
	CGCCTCGCCC	GCGATGTTTG	GCCAGCAGCT	GCGTCCGAC	GTGCAGCAGT	ACCTGGAGCG	8520
	CCTGGAGAAA	CAGAGGCAAC	AGAAGGTGGG	CCTCGACGAG	GCGTCGGCGG	GCCTGACGCT	8580
	CGGCGGCGAT	GCGCTGCGCG	TCCCTTTTTT	GGATTTTGCC	ACCGCGACGC	CCAAGCGCCA	8640
	CCAGACCGTG	GTCCCGGGCG	TCGGGACGCT	CCACGACTGC	TGCGAGCACT	CGCCGCTCTT	8700
25	CTCGGCCGTC	GCGCGGCGGT	TGCTGTTTAA	TAGCCTGGTG	CCGGCGCAAC	TCAGGGGGCG	8760
	TGACTTTGGG	GGCGACCACA	CGGCCAAGCT	GGAGTTCCCTG	GCCCCCGAGC	TGGTGCGGGC	8820
	GGTGGCGCGC	CTGCGGTTTC	GGGAGTGCGC	GCCGGAGGAC	GCCGTGCCCC	AACGCAACGC	8880
	CTACTACAGC	GTCCTGAACA	CGTTTCAGGC	CCTGCACCGC	TCCGAAGCCT	TTCGGCAGTT	8940
	GGTTCACTTC	GTGCGGGACT	TCGCCAGTT	GTGAAAACC	TCGTTCCGGG	CCTCTAGTCT	9000
30	CGCGGAGACT	ACGGGCCCCC	CGAAGAAACG	GGCCAAGGTG	GACGTGGCCA	CCCACGGGCA	9060
	GACGTACGGC	ACCTTGAGAG	TCTTCCAGAA	AATGATACTA	ATGCACGCGA	CCTACTTTCT	9120
	GGCCGCCGTG	CTGCTCGGGG	ACCACGCGGA	GCAGGTCAAC	ACGTTCCCTG	GGCTCGTGTT	9180
	CGAGATCCCC	CTGTTTAGCG	ACACGGCCGT	GCGGCACCTC	CGCCAGCGCG	CCACCGTGTT	9240
	TCTAGTCCCC	AGGCGCCACG	GAAAGACCTG	GTTTTTGGTG	CCCCTCATCG	CGCTGTCGCT	9300
35	CGCGTCCTTC	CGGGGGATCA	AGATAGGCTA	CACGGCCAC	ATCCGCAAGG	CGACCGAGCC	9360
	CGTGTGTTGAT	GAGATCGACG	CCTGCCTGCG	GGGCTGGTTT	GGCTCGTCCC	GGGTGGACCA	9420
	CGTCAAGGGG	GAAACCATCT	CGTTCTCGTT	CCCGGACGGC	TCGCGCAGCA	CGATCGTGTT	9480
	TGCCCTCCAGC	CACAACACGA	ACGTAAGTAC	GCCTTCCTCC	CGCGGTGCCT	GTTTCCCCCG	9540
	TGCCGCCCTC	CCCGAGATCG	ACCGACAGAC	AAACACAGCC	AGACGCGAGT	GTGGGACGAC	9600
40	ACGCCCGCAG	CCCCCCCCGC	CATGGCGGGG	GGAAGCCTTA	CTGTTTATTT	GTAATCGGAC	9660
	GATGAGGCTC	TGGCCACGGC	CCGCGCGACC	GCGGGGCAGC	TCGTTGCAAA	CAGGCGGCTG	9720
	GTATACGATG	ACAGAACGCA	GAGGCGCCAC	CCGGCGCTGG	TCGGGCGGAT	GACGCTTTCC	9780
	GCGCCGTCCC	GGCCACGAC	GACCTCGTGC	AGGTGGGCGG	TGATGCGCGG	GCGGCGGGTC	9840

GCCTGCCGCA GGATAACCGC GTCCACGGGG TCCCCGAAGA GGAGCTGACA CAGGCTCGCG 9900  
 TCCCCCGGA CGGCCAGGGT GCGCTGGGCC ATATTGGACC ACATGCACGG GGCGACGCAG 9960  
 GGACAGGCCT CCGCCACGGC GGGGGCGCGC CACAGCGCGT TGGCGGAATC GATGTGGGCC 10020  
 GTCGGGGCGC AGGCGCCGCC TCCTCCCGGG GGGTCGGTAA TCCTGGATAG CAGCCATCCT 10080  
 5 AAATGGCGGG CCCGGCTGCC CGGGGACAG AGCGACCCCA GGTCATCATC CATGGCCCAG 10140  
 CAGTATATGC GGCCGCCGGG GAGGTGCCAC CAGGCCCCCG GACCCAGGGC ACAGCACGCC 10200  
 CCCGGATTTC GGGGCGGTTC CGTGGGTACC AGGTAGGCGC CGTCGAGCTC GTGGGCCACG 10260  
 GGCTCGTCCG CGAGCTGTTC GCGGCGGGG TCGGGGGTTT CCTCCGGGGG GGAGGCAGCT 10320  
 TCCAGGTGGC CGAAGGCTAG GGTGCACAGC AGCGGGGTCC GGGGGTTCGT TACGCTGCGG 10380  
 10 AGGTGGACGG TGGCGCAGTA GCGGCGCTCG CGGTTAAAGA AGAAAATGGC AAAGAACGTG 10440  
 TTCGAAGGCA GCGCGAGCGC CTTGGGCCGC GTCAGGTACA GGAAGATCTC GCAGAAAAGG 10500  
 GCACGCTCGG GGTGCGGGTC CGGAAGGGCC ACCTGGCACA GCGGCTCGGT GAGGACCGTG 10560  
 AGGCACCGAA AAATCTTAAG CCGCTCGTCC CCCGAACGA CGCGCCACAC GAAGACAGAG 10620  
 TTGGCGATGC GCGCGACGAG GTCGGCTTCG GGCCCCGGGT CGGGGGCGCG CGCGTCGGGG 10680  
 15 GGGGCGCCCC GGTGACCCGG CGGGGCCGCG GCTCCCGGGG GGCTTGGCGT CGCCTGGGGA 10740  
 CGCCAGAGTG CCCGCTGTGC CAGGTTGGTG GTGGGGAAGG GACCGGAGAC GCACCAAAAG 10800  
 CAGAGGGGCC AGCGCGTGTG TGAGTTGGGG GGGGGGTGGG TGAGCGGTGG AACAAAAGCA 10860  
 CGCGTCAGCG GACAAGGCCG GGTCCCGTAG CCGCCCCGCG ACAGAACCGG AGTCCGACGG 10920  
 CACGCGCGAC GGGGTCTGCG AGGCTGAGGT ACGCCGCGGT GTTAATGGTA AACGCAAAGC 10980  
 20 CTCCCGGAAA GACCACTAGC CCGCAGAGGC GGCATTGAA CCCAAGGCAG AGGTACGCGT 11040  
 AGCTCTCTCC CGGAAGGTAT TGCTCGCAGA CCCTGTGCGG GGCAGTGGAG GGGCTGCCCT 11100  
 CCATGAAGCG ACATTTACTC TGCTCGCGTC CATTGACGTC ACCGTCAATC ACCACTGCGA 11160  
 TTGGACGGTT GGTAAGGCGC AGCGTGCTC CGCTGGTGCT GTAGTAGTCA AACCGTAGT 11220  
 GGGCGTCGGA GTCGGCGAAG CGGGCGGGGA TGTCGTGCT GAGAGGGACG AGCCGCCGCC 11280  
 25 GCCGCCCCCG ACCGCCCTGG CCGCCAGAT GCGCCAGCAC GGCCAGGGCG TACGCGGTGT 11340  
 GAAAGAACGC GTCGGGGGCG GTCCCTCGA GGGCGCGCAT CAGGTTCTCC AGGAGCACGG 11400  
 GGAAGCGCCG CGTCACCTCC CCTAGCCACT CGTCTGGTG GGGGCCAAAG TCGTAGCGCA 11460  
 GCGCTGGAA GATGCGCGG CCGCCTTGA GCGCGGCCG GATAGAGTGG CCCAGGGCCC 11520  
 GCAGACACGC GATCTGGATG CGCGCGACGA AGGCCACCTC GGCCGCGATG TCAAAGGGCT 11580  
 30 GCAGCACGGG GCGCGGGTGG CGCAGGGGTC CCTCGAGCGC GGGAAAGCGA CGCAGCAGCG 11640  
 CCGTCTGGGC CGCGGGGGAC AGCTGGTGGG GGCGCACGAC GCGCTCGGCG GCACAGGCCT 11700  
 CCGTC 11706

## (2) INFORMATION FOR SEQ ID NO:156:

35

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 63 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

40

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

```

Met Glu Ala Pro Gly Ile Val Trp Val Glu Glu Ser Val Ser Ala Ile
  1           5           10           15
5  Thr Leu Tyr Ala Val Trp Leu Pro Pro Arg Thr Arg Asp Cys Leu His
    20           25           30
Ala Leu Leu Tyr Leu Val Cys Arg Asp Ala Ala Gly Glu Ala Arg Ala
    35           40           45
Arg Phe Ala Glu Val Ser Val Gly Ser Ser Xaa Xaa Xaa Xaa Xaa
10  50           55           60

```

## (2) INFORMATION FOR SEQ ID NO:157:

## (i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 857 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

```

Met Ala Glu Thr Met Asn Val Ala Thr Cys Thr His Gln Thr His His
25  1           5           10           15
Ala Ala Arg Ala Pro Gly Ala Thr Ser Ala Pro Gly Ala Ala Ser Gly
    20           25           30
Asp Pro Leu Gly Ala Arg Arg Pro Ile Gly Asp Asp Glu Cys Glu Gln
    35           40           45
30  Tyr Thr Ser Ser Val Ser Leu Ala Arg Met Leu Tyr Gly Gly Asp Leu
    50           55           60
Ala Glu Trp Val Pro Arg Val His Pro Lys Thr Thr Ile Glu Arg Gln
    65           70           75           80
Gln His Gly Pro Val Thr Phe Pro Asp Ala Ser Ala Pro Thr Ala Arg
35  85           90           95
Cys Val Thr Val Val Arg Ala Pro Met Gly Ser Gly Lys Thr Thr Ala
    100          105          110
Leu Ile Arg Trp Leu Gly Glu Ala Ile His Ser Pro Asp Thr Ser Val
    115          120          125
40  Leu Val Val Ser Cys Arg Arg Ser Phe Thr Gln Thr Leu Ala Thr Arg
    130          135          140
Phe Ala Glu Ser Gly Leu Pro Asp Phe Val Thr Tyr Phe Ser Ser Thr
    145          150          155          160

```

Asn Tyr Ile Met Asn Asp Arg Pro Phe His Arg Leu Ile Val Gln Val  
 165 170 175  
 Glu Ser Leu His Arg Val Gly Pro Asn Leu Leu Asn Asn Tyr Asp Val  
 180 185 190  
 5 Leu Val Leu Asp Glu Val Met Ser Thr Leu Gly Gln Lys Pro Thr Met  
 195 200 205  
 Gln Gln Leu Gly Arg Val Asp Ala Leu Met Leu Arg Leu Leu Arg Thr  
 210 215 220  
 Cys Pro Arg Ile Ile Ala Met Asp Ala Thr Ala Asn Ala Gln Leu Val  
 10 225 230 235 240  
 Asp Phe Leu Cys Ser Leu Arg Gly Glu Lys Asn Val His Val Val Ile  
 245 250 255  
 Gly Glu Tyr Ala Met Pro Gly Phe Ser Ala Arg Arg Cys Leu Phe Leu  
 260 265 270  
 15 Pro Arg Leu Gly Pro Glu Val Leu Gln Ala Ala Leu Arg Pro Pro Gly  
 275 280 285  
 Pro Ala Gly Gly Ala Pro Pro Pro Asp Ala Pro Pro Asp Ala Thr Phe  
 290 295 300  
 Phe Gly Glu Leu Glu Ala Arg Leu Ala Gly Gly Asp Asn Val Cys Ile  
 20 305 310 315 320  
 Phe Ser Ser Thr Val Ser Phe Ala Glu Val Val Ala Arg Phe Cys Arg  
 325 330 335  
 Gln Phe Thr Asp Arg Val Leu Leu Leu His Ser Leu Thr Pro Pro Gly  
 340 345 350  
 25 Asp Val Thr Thr Trp Gly Arg Tyr Arg Val Val Ile Tyr Thr Thr Val  
 355 360 365  
 Val Thr Val Gly Leu Ser Phe Asp Pro Pro His Phe Asp Ser Met Phe  
 370 375 380  
 Ala Tyr Val Lys Pro Met Asn Tyr Gly Pro Asp Met Val Ser Val Tyr  
 30 385 390 395 400  
 Gln Ser Leu Gly Arg Val Arg Thr Leu Arg Lys Gly Glu Leu Leu Ile  
 405 410 415  
 Tyr Met Asp Gly Ser Gly Ala Arg Ser Glu Pro Val Phe Thr Pro Met  
 420 425 430  
 35 Leu Leu Asn His Val Val Ser Ala Ser Gly Gln Trp Pro Ala Gln Phe  
 435 440 445  
 Ser Gln Val Thr Asn Leu Leu Cys Arg Arg Phe Lys Gly Arg Cys Asp  
 450 455 460  
 Ala Ser His Ala Asp Ala Ala Gln Arg Ser Arg Ile Tyr Ser Lys Phe  
 40 465 470 475 480  
 Arg Tyr Lys His Tyr Phe Glu Arg Cys Thr Leu Ala Cys Leu Ala Asp  
 485 490 495  
 Ser Leu Asn Ile Leu His Met Leu Leu Thr Leu Asn Cys Met His Val  
 446

	500	505	510
	Arg Phe Trp Gly His Asp Ala Ala Leu Thr Pro Arg Asn Phe Cys Leu		
	515	520	525
	Phe Leu Arg Gly Ile His Phe Asp Ala Leu Arg Ala Gln Arg Asp Leu		
5	530	535	540
	Arg Glu Leu Arg Cys Gln Asp Pro Asp Thr Ser Leu Ser Ala Gln Ala		
	545	550	555
	Ala Glu Thr Glu Glu Val Gly Leu Phe Val Glu Lys Tyr Leu Arg Pro		
	565	570	575
10	Asp Val Ala Pro Ala Glu Val Val Met Arg Gln Ser Leu Val Gly Arg		
	580	585	590
	Thr Arg Phe Ile Tyr Leu Val Leu Leu Glu Ala Cys Leu Arg Val Pro		
	595	600	605
	Met Ala Ala His Ser Ser Ala Ile Phe Arg Arg Leu Tyr Asp His Tyr		
15	610	615	620
	Ala Thr Gly Val Ile Pro Thr Ile Asn Ala Ala Gly Glu Leu Glu Leu		
	625	630	635
	Val His Pro Thr Leu Asn Val Ala Pro Val Trp Glu Leu Phe Arg Leu		
	645	650	655
20	Cys Ser Thr Met Ala Ala Cys Leu Gln Trp Asp Ser Met Ala Gly Gly		
	660	665	670
	Ser Gly Arg Thr Phe Ser Pro Glu Asp Val Leu Glu Leu Leu Asn Pro		
	675	680	685
	His Tyr Asp Arg Tyr Met Gln Leu Val Phe Glu Leu Gly His Cys Asn		
25	690	695	700
	Val Thr Asp Gly Pro Leu Leu Ser Glu Asp Ala Val Lys Arg Val Ala		
	705	710	715
	Asp Ala Leu Ser Gly Cys Pro Pro Arg Gly Ser Val Ser Glu Thr Glu		
	725	730	735
30	His Ala Leu Ser Leu Phe Lys Ile Ile Trp Gly Glu Leu Phe Gly Val		
	740	745	750
	Gln Leu Ala Lys Ser Thr Gln Thr Phe Pro Gly Ala Gly Arg Val Lys		
	755	760	765
	Asn Leu Thr Lys Arg Ala Ile Val Glu Leu Leu Asp Ala His Arg Ile		
35	770	775	780
	Asp His Ser Ala Cys Arg Thr Gln Leu Tyr Ala Leu Leu Met Ala His		
	785	790	795
	Lys Arg Glu Phe Ala Gly Ala Arg Phe Lys Leu Arg Ala Pro Ala Trp		
	805	810	815
40	Gly Arg Cys Leu Arg Thr His Ala Ser Gly Ala Gln Pro Asn Thr Asp		
	820	825	830
	Ile Ile Ala Ala Leu Ser Glu Leu Pro Thr Glu Ala Trp Pro Met Met		
	835	840	845

Gln Gly Ala Val Asn Phe Ser Thr Leu  
850 855

(2) INFORMATION FOR SEQ ID NO:158:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 470 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

10

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

15

Val	Tyr	Cys	Ser	His	Ser	Ser	Ser	Pro	Met	Gly	Arg	Arg	Ala	Pro	Arg
1				5					10					15	

Gly Ser Pro Glu Ala Ala Pro Gly Ala Asp Val Ala Pro Gly Ala Arg  
20 25 30

20 Ala Ala Trp Trp Val Trp Cys Val Gln Val Ala Thr Phe Ile Val Ser  
35 40 45

Ala Ile Cys Val Val Gly Leu Leu Val Leu Ala Ser Val Phe Arg Asp  
50 55 60

Arg Phe Pro Cys Leu Tyr Ala Pro Ala Thr Ser Tyr Ala Glu Ala Asn  
25      65                          70                          75                          80

Ala Thr Val Glu Val Arg Gly Gly Val Ala Val Pro Leu Arg Leu Asp  
85 90 95

Thr Gln Ser Leu Leu Ala Thr Tyr Ala Ile Thr Ser Thr Leu Leu Leu  
100 105 110

30 Ala Ala Ala Val Tyr Ala Ala Val Gly Ala Val Thr Ser Arg Tyr Glu  
115 120 125

Arg Ala Leu Asp Ala Ala Arg Arg Leu Ala Ala Ala Arg Met Ala Met  
130 135 140

	Pro	His	Ala	Thr	Leu	Ile	Ala	Gly	Asn	Val	Cys	Ala	Trp	Leu	Leu	Gln
35	145					150					155					160

Ile Thr Val Leu Leu Leu Ala His Arg Ile Ser Gln Leu Ala His Leu  
165 170 175

Ile Tyr Val Leu His Phe Ala Cys Leu Val Tyr Leu Ala Ala His Phe  
180 185 190

40 Cys Thr Arg Gly Val Leu Ser Gly Thr Tyr Leu Arg Gln Val His Gly  
195 200 205

Leu Ile Asp Pro Ala Pro Thr His His Arg Ile Val Gly Pro Val Arg  
210 215 220

Ala Val Met Thr Asn Ala Leu Leu Leu Gly Thr Leu Leu Cys Thr Ala  
 225 230 235 240  
 Ala Ala Ala Val Ser Leu Asn Thr Ile Ala Ala Leu Asn Phe Asn Phe  
 245 250 255  
 5 Ser Ala Pro Ser Met Leu Ile Cys Leu Thr Thr Leu Phe Ala Leu Leu  
 260 265 270  
 Val Val Ser Leu Leu Leu Val Val Glu Gly Val Leu Cys His Tyr Val  
 275 280 285  
 Arg Val Leu Val Gly Pro His Leu Gly Ala Ile Ala Ala Thr Gly Ile  
 10 290 295 300  
 Val Gly Leu Ala Cys Glu His Tyr His Thr Gly Gly Tyr Tyr Val Val  
 305 310 315 320  
 Glu Gln Gln Trp Pro Gly Ala Gln Thr Gly Val Arg Val Val Ala Ala  
 325 330 335  
 15 Phe Ala Met Ala Val Leu Arg Cys Thr Arg Ala Tyr Leu Tyr His Arg  
 340 345 350  
 Arg His His Thr Lys Phe Phe Val Arg Met Arg Asp Thr Arg His Arg  
 355 360 365  
 Ala His Ser Ala Leu Arg Arg Val Arg Ser Ser Met Arg Gly Ser Arg  
 20 370 375 380  
 Arg Gly Gly Pro Pro Gly Asp Pro Gly Tyr Ala Glu Thr Pro Tyr Ala  
 385 390 395 400  
 Ser Val Ser His His Ala Glu Ile Asp Arg Tyr Gly Asp Ser Asp Gly  
 405 410 415  
 25 Asp Pro Ile Tyr Asp Glu Val Ala Pro Asp His Glu Ala Glu Leu Tyr  
 420 425 430  
 Ala Arg Val Gln Arg Pro Gly Pro Val Pro Asp Ala Glu Pro Ile Tyr  
 435 440 445  
 Asp Thr Val Glu Gly Tyr Ala Pro Arg Ser Ala Gly Glu Pro Val Tyr  
 30 450 455 460  
 Ser Thr Val Arg Arg Trp  
 465 470

## (2) INFORMATION FOR SEQ ID NO:159:

35

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 96 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

40

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide



## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

Met Gly Leu Ala Phe Ser Gly Ala Arg Pro Cys Cys Cys Arg His Asn  
 1 5 10 15  
 5 Val Ile Ile Thr Asp Gly Gly Glu Val Val Ser Leu Thr Ala His Glu  
 20 25 30  
 Phe Asp Val Val Asp Ile Glu Ser Glu Glu Glu Gly Asn Phe Tyr Val  
 35 40 45  
 Pro Pro Asp Met Arg Val Val Thr Arg Ala Pro Gly Pro Gln Tyr Arg  
 10 50 55 60  
 Arg Ala Ser Asp Pro Pro Ser Arg His Thr Arg Arg Arg Asp Pro Asp  
 65 70 75 80  
 Val Ala Arg Pro Pro Ala Thr Leu Thr Pro Pro Leu Ser Asp Ser Glu  
 85 90 95

15

## (2) INFORMATION FOR SEQ ID NO:160:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 618 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20

## (ii) MOLECULE TYPE: peptide

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

Met Ala Ala Ala Ala Thr Pro Gly Ala Lys Arg Pro Ala Asp Pro Ala  
 1 5 10 15  
 30 Arg Asp Pro Asp Ser Pro Pro Lys Arg Pro Arg Pro Asn Ser Leu Asp  
 20 25 30  
 Leu Ala Thr Val Phe Gly Pro Arg Pro Ala Pro Pro Arg Pro Thr Ser  
 35 40 45  
 Pro Gly Ala Pro Gly Ser His Trp Pro Gln Ser Pro Pro Arg Gly Gln  
 50 55 60  
 Pro Asp Gly Gly Ala Pro Gly Glu Lys Ala Arg Pro Asp Ala Leu Ser  
 65 70 75 80  
 Glu Ala Ser Ser Gly Pro Pro Thr Pro Asp Ile Pro Leu Ser Pro Gly  
 85 90 95  
 40 Gly Ala His Ala Ile Asp Pro Asp Cys Ser Pro Gly Pro Pro Asp Pro  
 100 105 110  
 Asp Pro Met Trp Ser Ala Ser Ala Ile Pro Asn Ala Leu Pro Pro His  
 115 120 125

Ile Leu Ala Glu Thr Phe Glu Arg His Leu Arg Gly Leu Leu Arg Gly  
 130 135 140  
 Val Arg Ser Pro Leu Ala Ile-Gly Pro Leu Trp Ala Arg Leu Asp Tyr  
 145 150 155 160  
 5 Leu Cys Ser Leu Val Val Ser Leu Glu Ala Ala Gly Met Val Asp Arg  
 165 170 175  
 Gly Leu Gly Arg His Leu Trp Arg Leu Thr Arg Arg Ala Pro Pro Ser  
 180 185 190  
 Ala Ala Glu Ala Val Ala Pro Arg Pro Leu Met Gly Phe Tyr Glu Ala  
 10 195 200 205  
 Ala Thr Gln Asn Gln Ala Asp Cys Gln Leu Trp Ala Leu Leu Arg Arg  
 210 215 220  
 Gly Leu Thr Thr Ala Ser Thr Leu Arg Trp Gly Ala Gln Gly Pro Cys  
 225 230 235 240  
 15 Phe Ser Ser Gln Trp Leu Thr His Asn Ala Ser Leu Arg Leu Asp Ala  
 245 250 255  
 Gln Ser Ser Ala Val Met Phe Gly Arg Val Asn Glu Pro Thr Ala Arg  
 260 265 270  
 Asn Leu Leu Phe Arg Tyr Cys Val Gly Arg Ala Asp Ala Gly Val Asn  
 20 275 280 285  
 Asp Asp Ala Asp Ala Gly Arg Phe Val Phe His Gln Pro Gly Asp Leu  
 290 295 300  
 Ala Glu Glu Asn Val His Ala Cys Gly Val Leu Met Asp Gly His Thr  
 305 310 315 320  
 25 Gly Met Val Gly Ala Ser Leu Asp Ile Leu Val Cys Pro Arg Asp Pro  
 325 330 335  
 His Gly Tyr Leu Ala Pro Ala Pro Gln Thr Pro Leu Ala Phe Tyr Glu  
 340 345 350  
 Val Lys Cys Arg Ala Lys Tyr Ala Phe Asp Pro Ala Asp Pro Gly Ala  
 30 355 360 365  
 Pro Ala Ala Ser Ala Tyr Glu Asp Leu Met Ala Arg Arg Ser Pro Glu  
 370 375 380  
 Ala Phe Arg Ala Phe Ile Arg Ser Ile Pro Asn Pro Gly Val Arg Tyr  
 385 390 395 400  
 35 Phe Ala Pro Gly Arg Val Pro Gly Pro Glu Glu Ala Leu Val Thr Gln  
 405 410 415  
 Asp Arg Asp Trp Leu Asp Ser Arg Ala Ala Gly Glu Lys Arg Arg Cys  
 420 425 430  
 Ser Ala Pro Asp Arg Ala Leu Val Glu Leu Asn Ser Gly Val Val Ser  
 40 435 440 445  
 Glu Val Leu Leu Phe Gly Val Pro Asp Leu Glu Arg Arg Thr Ile Ser  
 450 455 460  
 Pro Val Ala Trp Ser Ser Gly Glu Leu Val Arg Arg Glu Pro Ile Phe  
 451

465                      470                      475                      480  
 Ala Asn Pro Arg His Pro Asn Phe Lys Gln Ile Leu Val Gln Gly Tyr  
                                  485                      490                      495  
 Val Leu Asp Ser His Phe Pro Asp Cys Pro Leu Gln Pro His Leu Val  
 5                      500                      505                      510  
 Thr Phe Leu Gly Arg His Arg Ala Gly Ala Glu Glu Gly Val Thr Phe  
                                  515                      520                      525  
 Arg Leu Glu Asp Gly Arg Gly Ala Pro Ala Gly Arg Gly Gly Ala Pro  
                                  530                      535                      540  
 10 Gly Pro Ala Lys Ala Ser Ile Leu Pro Asp Gln Ala Val Pro Ile Ala  
                                  545                      550                      555                      560  
 Leu Ile Ile Thr Pro Val Arg Val Glu Pro Gly Ile Tyr Arg Asp Ile  
                                  565                      570                      575  
 Arg Arg Asn Ser Arg Leu Ala Phe Asp Asp Thr Leu Ala Lys Leu Trp  
 15                      580                      585                      590  
 Ala Ser Arg Ser Pro Gly Arg Gly Pro Ala Ala Ala Asp Thr Thr Ser  
                                  595                      600                      605  
 Ser Ser Pro Thr Ala Gly Arg Ser Ser Arg  
                                  610                      615

20

(2) INFORMATION FOR SEQ ID NO:161:

(i) SEQUENCE CHARACTERISTICS:

25

- (A) LENGTH: 525 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

Val Gly Gly Arg Arg Pro Gly Gly Arg Met Asp Glu Ser Gly Arg Gln  
   1                      5                      10                      15  
 35 Arg Pro Ala Ser His Val Ala Ala Asp Ile Ser Pro Gln Gly Ala His  
                                  20                      25                      30  
 Arg Arg Ser Phe Lys Ala Trp Leu Ala Ser Tyr Ile His Ser Leu Ser  
                                  35                      40                      45  
 Arg Arg Ala Ser Gly Arg Pro Ser Gly Pro Ser Pro Arg Asp Gly Ala  
 40                      50                      55                      60  
 Val Ser Gly Ala Arg Pro Gly Ser Arg Arg Arg Ser Ser Phe Arg Glu  
   65                      70                      75                      80  
 Arg Leu Arg Ala Gly Leu Ser Arg Trp Arg Val Ser Arg Ser Ser Arg

				85					90				95						
				Arg	Arg	Ser	Ser	Pro	Glu	Ala	Pro	Gly	Pro	Ala	Ala	Lys	Leu	Arg	Arg
				100					105				110						
				Pro	Pro	Leu	Arg	Arg	Ser	Glu	Thr	Ala	Met	Thr	Ser	Pro	Pro	Ser	Pro
5				115					120				125						
				Pro	Ser	His	Ile	Leu	Ser	Leu	Ala	Arg	Ile	His	Lys	Leu	Cys	Ile	Pro
				130					135				140						
				Val	Phe	Ala	Val	Asn	Pro	Ala	Leu	Arg	Tyr	Thr	Thr	Leu	Glu	Ile	Pro
				145					150				155					160	
10				Gly	Ala	Arg	Ser	Phe	Gly	Gly	Ser	Gly	Gly	Tyr	Gly	Glu	Val	Gln	Leu
								165				170					175		
				Ile	Arg	Glu	His	Lys	Leu	Ala	Val	Lys	Thr	Ile	Arg	Glu	Lys	Glu	Trp
								180				185					190		
				Phe	Ala	Val	Glu	Leu	Val	Ala	Thr	Leu	Leu	Val	Gly	Glu	Cys	Ala	Leu
15				195					200				205						
				Arg	Gly	Gly	Arg	Thr	His	Asp	Ile	Arg	Gly	Phe	Ile	Thr	Pro	Leu	Gly
				210					215				220						
				Phe	Ser	Leu	Gln	Gln	Arg	Gln	Ile	Val	Phe	Pro	Ala	Tyr	Asp	Met	Asp
				225					230				235					240	
20				Leu	Gly	Lys	Tyr	Ile	Gly	Gln	Leu	Ala	Ser	Leu	Arg	Ala	Thr	Thr	Pro
								245				250					255		
				Ser	Val	Ala	Thr	Ala	Leu	His	His	Cys	Phe	Thr	Asp	Leu	Ala	Arg	Ala
								260				265					270		
				Val	Val	Phe	Leu	Asn	Thr	Arg	Cys	Gly	Ile	Ser	His	Leu	Asp	Ile	Lys
25				275					280				285						
				Cys	Ala	Asn	Val	Leu	Val	Met	Leu	Arg	Ser	Asp	Ala	Val	Ser	Leu	Arg
				290					295				300						
				Arg	Ala	Val	Leu	Ala	Asp	Phe	Ser	Leu	Val	Thr	Leu	Asn	Ser	Asn	Ser
				305					310				315					320	
30				Thr	Ile	Ser	Arg	Gly	Gln	Phe	Cys	Leu	Gln	Glu	Pro	Asp	Leu	Glu	Ser
								325				330					335		
				Pro	Arg	Gly	Phe	Gly	Met	Pro	Ala	Ala	Leu	Thr	Thr	Ala	Asn	Phe	His
								340				345					350		
				Thr	Leu	Val	Gly	His	Gly	Tyr	Asn	Gln	Pro	Pro	Glu	Leu	Leu	Val	Lys
35				355					360				365						
				Tyr	Leu	Asn	Asn	Glu	Arg	Ala	Glu	Phe	Asn	Asn	Arg	Pro	Leu	Lys	His
				370					375				380						
				Asp	Val	Gly	Leu	Ala	Val	Asp	Leu	Tyr	Ala	Leu	Gly	Gln	Thr	Leu	Leu
				385					390				395					400	
40				Glu	Leu	Leu	Val	Ser	Val	Tyr	Val	Ala	Pro	Ser	Leu	Gly	Val	Pro	Val
								405				410					415		
				Thr	Arg	Val	Pro	Gly	Tyr	Gln	Tyr	Phe	Asn	Asn	Gln	Leu	Ser	Pro	Asp
								420				425					430		

Phe Ala Val Leu Ala Tyr Arg Cys Val Leu His Pro Ala Leu Phe Val  
           435                          440                          445  
 Asn Ser Ala Glu Thr Asn Thr His Gly Leu Ala Tyr Asp Val Pro Glu  
           450                          455                          460  
 5 Gly Ile Arg Arg His Leu Arg Asn Pro Lys Ile Arg Arg Ala Phe Thr  
    465                          470                          475                          480  
 Glu Gln Cys Ile Asn Tyr Gln Arg Thr His Lys Ala Val Leu Ser Ser  
                           485                          490                          495  
 Val Ser Leu Pro Pro Glu Leu Arg Pro Leu Leu Val Leu Val Ser Arg  
 10                          500                          505                          510  
 Leu Cys His Ala Asn Pro Ala Ala Arg His Ser Leu Ser  
           515                          520                          525

## (2) INFORMATION FOR SEQ ID NO:162:

15

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 217 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 20 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

25

Met Ser Arg Asp Ala Ser His Ala Ala Leu Arg Arg Arg Leu Ala Glu  
    1                          5                          10                          15  
 Thr His Leu Arg Ala Glu Val Tyr Arg Asp Gln Thr Leu Gln Leu His  
                           20                          25                          30  
 30 Arg Glu Gly Val Ser Thr Gln Asp Pro Arg Phe Val Gly Ala Phe Met  
           35                          40                          45  
 Ala Ala Lys Ala Ala His Leu Glu Leu Glu Ala Arg Leu Lys Ser Arg  
           50                          55                          60  
 Ala Arg Leu Glu Met Met Arg Gln Arg Ala Thr Cys Val Lys Ile Arg  
 35 65                          70                          75                          80  
 Val Glu Glu Gln Ala Ala Arg Arg Asp Phe Leu Thr Ala His Arg Arg  
                           85                          90                          95  
 Tyr Leu Asp Pro Ala Leu Ser Leu Asp Ala Ala Asp Asp Arg Leu Ala  
           100                          105                          110  
 40 Asp Gln Glu Glu Gln Leu Glu Glu Ala Ala Ala Asn Ala Ser Leu Trp  
           115                          120                          125  
 Gly Asp Gly Asp Leu Ala Asp Gly Trp Met Ser Pro Gly Asp Ser Asp  
           130                          135                          140

Leu Leu Val Met Trp Gln Leu Thr Ser Ala Pro Lys Val His Thr Asp  
 145 150 155 160  
 Ala Pro Ser Arg Pro Gly Ser-Arg Pro Thr Tyr Thr Pro Ser Ala Ala  
 165 170 175  
 5 Gly Arg Pro Asp Ala Gln Ala Ala Pro Pro Glu Thr Ala Pro Ser  
 180 185 190  
 Pro Glu Pro Ala Pro Gly Pro Ala Ala Asp Pro Ala Ser Gly Ser Gly  
 195 200 205  
 Phe Ala Arg Asp Cys Pro Asp Gly Glu  
 10 210 215

## (2) INFORMATION FOR SEQ ID NO:163:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 239 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

Pro Ala Asp Leu Glu Pro Leu Gly Asp Pro Thr Leu Trp Arg Ala Leu  
 25 1 5 10 15  
 Tyr Ala Cys Val Leu Ala Ala Leu Glu Arg Gln Thr Gly Pro Val Phe  
 20 25 30  
 Val Pro Leu Arg Leu Gly Trp Asp Pro Gln Thr Gly Leu Val Val Arg  
 35 40 45  
 30 Val Glu Arg Ala Ser Trp Gly Pro Pro Ala Ala Pro Arg Ala Ala Leu  
 50 55 60  
 Leu Asp Val Glu Ala Lys Val Asp Val Asp Pro Leu Ala Ala Arg Val  
 65 70 75 80  
 Ala Glu His Pro Gly Ala Arg Leu Ala Trp Ala Arg Leu Ala Ala Ile  
 35 85 90 95  
 Arg Asp Ser Pro Gln Cys Ala Ser Ser Ala Ser Leu Ala Val Thr Ile  
 100 105 110  
 Thr Thr Arg Thr Ala Arg Phe Ala Arg Glu Tyr Thr Thr Leu Ala Phe  
 115 120 125  
 40 Pro Pro Thr Ser Lys Glu Gly Ala Phe Ala Asp Leu Val Glu Val Cys  
 130 135 140  
 Glu Val Gly Leu Arg Pro Arg Gly His Pro Gln Arg Val Thr Ala Arg  
 145 150 155 160

Val Leu Leu Pro Arg Gly Tyr Asp Tyr Phe Val Ser Ala Gly Asp Gly  
 165 170 175  
 Phe Ser Ala Pro Ala Leu Val Phe Arg Gln Trp His Thr Thr Val His  
 180 185 190  
 5 Ala Ala Pro Gly Ala Pro Val Phe Ala Phe Leu Gly Ala Gly Phe Asp  
 195 200 205  
 Val Arg Gly Gly Pro Val Gln Tyr Phe Ala Val Leu Gly Phe Pro Gly  
 210 215 220  
 Trp Pro Thr Phe Thr Val Pro Ala Ala Ala Xaa Xaa Xaa Xaa Xaa  
 10 225 230 235

## (2) INFORMATION FOR SEQ ID NO:164:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 315 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

Val Trp Arg Val Val Arg Gly Asp Glu Arg Leu Lys Ile Phe Arg Cys  
 25 1 5 10 15  
 Leu Thr Val Leu Thr Glu Pro Leu Cys Gln Val Pro Asp Pro Asp Pro  
 20 25 30  
 Glu Arg Ala Leu Phe Cys Glu Ile Phe Leu Tyr Leu Trp Lys Ala Leu  
 35 40 45  
 30 Arg Leu Pro Ser Asn Thr Phe Phe Ala Ile Phe Phe Phe Asn Arg Glu  
 50 55 60  
 Arg Arg Tyr Cys Ala Thr Val His Leu Arg Ser Val Thr His Pro Arg  
 65 70 75 80  
 Thr Pro Leu Leu Cys Thr Leu Ala Phe Gly His Leu Glu Ala Asp Pro  
 35 85 90 95  
 Glu Glu Thr Pro Asp Pro Ala Ala Glu Gln Leu Ala Asp Glu Pro Val  
 100 105 110  
 Ala His Glu Leu Asp Gly Ala Tyr Leu Val Pro Thr Glu Pro Pro Pro  
 115 120 125  
 40 Asn Pro Gly Ala Cys Cys Ala Leu Gly Pro Gly Ala Trp Trp His Leu  
 130 135 140  
 Pro Gly Gly Arg Ile Tyr Cys Trp Ala Met Asp Asp Asp Leu Gly Ser  
 145 150 155 160

Leu Cys Pro Pro Gly Ser Arg Ala Arg His Leu Gly Trp Leu Leu Ser  
 165 170 175  
 Arg Ile Thr Asp Pro Pro Gly Gly Gly Ala Cys Ala Pro Thr Ala  
 180 185 190  
 5 His Ile Asp Ser Ala Asn Ala Leu Trp Arg Ala Pro Ala Val Ala Glu  
 195 200 205  
 Ala Cys Pro Cys Val Ala Pro Cys Met Trp Ser Asn Met Ala Gln Arg  
 210 215 220  
 Thr Leu Ala Val Arg Gly Asp Ala Ser Leu Cys Gln Leu Leu Phe Gly  
 10 225 230 235 240  
 His Pro Val Asp Ala Val Ile Leu Arg Gln Ala Thr Arg Arg Pro Arg  
 245 250 255  
 Ile Thr Ala His Leu His Glu Val Val Val Gly Arg Asp Gly Ala Glu  
 260 265 270  
 15 Ser Val Ile Arg Pro Thr Ser Ala Gly Trp Arg Leu Cys Val Leu Ser  
 275 280 285  
 Ser Tyr Thr Ser Arg Leu Phe Ala Thr Ser Cys Pro Ala Val Ala Arg  
 290 295 300  
 Ala Val Ala Arg Ala Ser Ser Ser Asp Tyr Lys  
 20 305 310 315

## (2) INFORMATION FOR SEQ ID NO:165:

## (i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 278 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 30 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

Leu Thr Glu Ala Cys Ala Ala Glu Arg Val Val Arg Pro His Gln Leu  
 35 1 5 10 15  
 Ser Pro Ala Ala Gln Thr Ala Leu Leu Arg Arg Phe Pro Ala Leu Glu  
 20 25 30  
 Gly Pro Leu Arg His Pro Arg Pro Val Leu Gln Pro Phe Asp Ile Ala  
 35 40 45  
 40 Ala Glu Val Ala Phe Val Ala Arg Ile Gln Ile Ala Cys Leu Arg Ala  
 50 55 60  
 Leu Gly His Ser Ile Arg Ala Ala Leu Gln Gly Gly Pro Arg Ile Phe  
 65 70 75 80



(2) INFORMATION FOR SEQ ID NO:166:

35

40	GCGGCGGCCG	GCACGGTAAA	CGTGGGCCAG	CCCGGAAATC	CCAGCACGGC	AAAGTATTGG	60
	ACGGGCCCTC	CCCGGACGTC	AAACCCGGCC	CCCAGAAAAG	CGAAGACGGG	GGCCAGGGCT	120
	CCGGGGGCGG	CGTGGACCGT	GGTATGCCAC	TGCCGGAAAG	GGGCGACCAG	CGCCGGGGCG	180
	GAGAACCCTG	CGCCGGCGCT	CACGAAAGTAG	TCGTAGCCGC	CGGGCAGCAG	CACCCGCGCC	240
	GTGACCCGCT	GCGGGTGTC	GCGGGGCCGC	AGGCCGACCT	CGCACACCTC	GACCAGGTCC	300

GCGAAGGCGC CCTCCTTGCT GGTGGGCGGA AACGCCAGGG TGGTGTATTC GCGCGCGAAA 360  
 CGCGCGGTCC TCGTCGTGAT GGTGACGGCG AGCGAGGCGG AGGACGCGCA CTGGGGGCTG 420  
 TCGCGAATGG CGGCCAGGCG CGCCCACGCC AACCGCGCGC CGGGGTGCTC GGCGACGCGC 480  
 GCGGCCAGGG CCAGCGGGTC GACGTCGACC TTGGCCTCCA CGTCCAGGAG GGCGGCGCGA 540  
 5 GGAGCGGCCG GCGGGCCCCA CGACGCCCTT TCGACCCTCA CGACCAGACC CGTCTGCGGG 600  
 TCCAGCCCCA GCGCGAGCGG GACGAAGAGG GCCACCGGCC CCGTCTGGCG CTCCAGGGCC 660  
 GCCAGAACGC ACGCATACAG CGCCCGCCAC AGGGTCGGGT CCCCAGGGG CTCCAGCGGG 720  
 GAGGCGGCCG 731

10 (2) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 239 amino acids  
 (B) TYPE: amino acid  
 15 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

20 Pro Ala Asp Leu Glu Pro Leu Gly Asp Pro Thr Leu Trp Arg Ala Leu  
 1 5 10 15  
 Tyr Ala Cys Val Leu Ala Ala Leu Glu Arg Gln Thr Gly Pro Val Phe  
 20 25 30  
 25 Val Pro Leu Arg Leu Gly Trp Asp Pro Gln Thr Gly Leu Val Val Arg  
 35 40 45  
 Val Glu Arg Ala Ser Trp Gly Pro Pro Ala Ala Pro Arg Ala Ala Leu  
 50 55 60  
 Leu Asp Val Glu Ala Lys Val Asp Val Asp Pro Leu Ala Ala Arg Val  
 30 65 70 75 80  
 Ala Glu His Pro Gly Ala Arg Leu Ala Trp Ala Arg Leu Ala Ala Ile  
 85 90 95  
 Arg Asp Ser Pro Gln Cys Ala Ser Ser Ala Ser Leu Ala Val Thr Ile  
 100 105 110  
 35 Thr Thr Arg Thr Ala Arg Phe Ala Arg Glu Tyr Thr Thr Leu Ala Phe  
 115 120 125  
 Pro Pro Thr Ser Lys Glu Gly Ala Phe Ala Asp Leu Val Glu Val Cys  
 130 135 140  
 Glu Val Gly Leu Arg Pro Arg Gly His Pro Gln Arg Val Thr Ala Arg  
 40 145 150 155 160  
 Val Leu Leu Pro Arg Gly Tyr Asp Tyr Phe Val Ser Ala Gly Asp Gly  
 165 170 175  
 Phe Ser Ala Pro Ala Leu Val Phe Arg Gln Trp His Thr Thr Val His

	180	185	190
	Ala Ala Pro Gly Ala Pro Val Phe Ala Phe Leu Gly Ala Gly Phe Asp		
	195	200	205
	Val Arg Gly Gly Pro Val Gln Tyr Phe Ala Val Leu Gly Phe Pro Gly		
5	210	215	220
	Trp Pro Thr Phe Thr Val Pro Ala Ala Ala Xaa Xaa Xaa Xaa		
	225	230	235

## (2) INFORMATION FOR SEQ ID NO:168:

10

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3005 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

15

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

20	GCGCTGGAGC GGGAGCAGCG CGCGGCGGAC CGGGCGGCCG GGGGAGGCGC GGGCCGCCCG	60
	GCGGAGGCGG ATCTTCTCCG GGCCGACTAC GACATTATCG ACGTCAGCAA GTCCATGGAC	120
	GACGACACGT ACGTGGCCAA CAGTTTCCAG CACCAGTACA TCCCCGCGTA CGGCCAGGAC	180
	CTCGAGCGCC TGTCGCGCCT CTGGGAGCAC GAGCTGGTGC GCTGCTTCAA GATTCTGCGC	240
	CACCGCAACA ACCAGGGCCA GGAAACGTCG ATCTCGTACT CTAGCGGGGC GATCGCCTCC	300
25	TTCGTGGCCC CGTATTTCGA GTACGTGCTT CGCGCCCCC GAGCGGGCGC GCTCATCACC	360
	GGCTCCGATG TCATCCTAGG GGAGGAGGAG TTATGGGAGG CGGTCTTTAA GAAAACCCGC	420
	CTGCAGACGT ACCTGACAGA CGTCGCGGCC CTGTTCTGTG CGGACGTACA GCACGCGGCT	480
	CTGCCCCGGC CCCCCTCCCC AACCCCCGCC GATTTCCGGG CGAGCGCGTC CCCGCGGGGC	540
	GGGTCCCGGT CCCGGACCCG GACCCGATCC CGGTCGCCCC GGAGAACGCC GAGGGGTGCG	600
30	CCGGAACAGG GCTGGGGCGT CGAACGCAGG GATGGCCGAC CCCACGCCCC CCGATGAGGG	660
	AACGGCCGCC GCCATCCTCA AACAGGCCAT CGCCGGGGAC CGCAGTCTGG TCGAGGTGGC	720
	GGAGGGGATC AGCAACCAGG CGCTGCTGCG CATGGCCTGC GAGGTGCGCC AGGTCAGCGA	780
	TCGCCAGCCG CGGTTTACCG CGACCAGCGT CCTGCGCGTT GACGTCACCC CCAGGGGGCG	840
	GTTGCGGTTC GTTCTGGACG GGAGTTCCGA CGACGCGTAC GTGGCGTCCG AGGATTACTT	900
35	TAAGCGCTGC GGGGACCAGC CGACGTATCG CGGTTTTCG GTCTGCTGCC TCACGGCCAA	960
	CGAGGACCAC GTGCACAGCC TGGCCGTGCC CCCCCTCGTT CTGCTGCACC GGCTCTCCTT	1020
	GTTTCGCCCC ACGGACCTCC GGGACTTCGA GCTCGTCTGC CTGCTGATGT ACCTGGAGAA	1080
	CTGTCCCCGG AGCCACGCCA CGCCCTCGCT GTTCGTCAAG GTGTCGGCGT GGTGGGGGT	1140
	CGTGGCCCCG CACGCGTCTC CCTTCGAGCG CGTCCGCTGC CTTCTCCTCC GCAGCTGCCA	1200
40	CTGGATCCTG AACACGCTAA TGTGCATGGC GGGCGTGAAG CCCTTCGACG ACGAGCTAGT	1260
	CCTGCCCCAC TGGTACATGG CCCACTACCT GCTGGCCAAC AATCCGCCCC CCGTCCCTCTC	1320
	GGCCCTGTTT TGCGCCACCC CGCAGAGCTC TGCCTGTCAG TTGCCCCGGC CCGTCCCCCG	1380
	CACGGACTGT GTGGCCTATA ACCCGGCCGG CGTCATGGGA AGCTGCTGGA AATCCAAGGA	1440

CCTGCGTTTCG GCTCTGGTGT ATTGGTGGCT TTCGGGGAGC CCCAAACGAC GGACCTCGTC 1500  
 GCTTTTCTAT CGGTTTTGCT AACTCCGGAA AATAAACGTG TTTTATATGG AACGTTCCCT 1560  
 ACCTGTCGTG TCATCTCTCG GGGGATGGTG GTGGGCCTGT GTGTGTGTCT TGTGCACCGA 1620  
 AGGAGGAAAG TGGGGGGGTG GTGGTGCTGG TGGTGGAAAG ACATGATAGA GGAACAAAG 1680  
 5 AAATAGAAGA AAACCACAAC CGGCGCGTGT CAGTAAATAC GGACGCGCGC ACACGCGGGG 1740  
 GTAAGTTGGA GCACGGGGCC CCAGTTTATT GACCAAATTC AGGGAAACAG AAACCGCATC 1800  
 TTTTCCTCGA AAGGGTACAC AAAGCTCCCG CCCTCGCCCC ACACGCCTTC CAGAACCCCC 1860  
 GTAAACACCA GTTGAATCTC GCGCAGGATC TCGCGCAGGT GATGGGCGCA GTCCACGGGG 1920  
 GGGAGCACCA AGGGCCGCGG GTACAGATCC ACGGGGACGC CGACCGACTC CCCGCCCCCG 1980  
 10 GGACATACGC GCACGACGCG TCTCCAGTAT TGCTCCGCGT CCAGCAGGGG GCCTCCGCGG 2040  
 AAGGCCGTTT GGGGCAGGGG GTCGTCGGCC TCGCCCGGGG GGGTCAGAAC GCTCCAGTAC 2100  
 TCCGCGTCCA GACGCCTCCC GAAGGCATCC AGGACAAAGC GGTACAGGC GTCCTCCATG 2160  
 ATGCCCCGGG CCGCGCACAC GGCCTCCTCC GCGGGGCGCG CGGCCGCGCG CCGGAGGATT 2220  
 CGTCTCAGCG CGTCGCGCAT AACCTCGGCC GCCGCGGCGT ACGCGGGCCC GCGGAGAGGA 2280  
 15 AATCCCTGCA GGAAGTCGGT GTCATCGCGG GAGTTCAGA ACCACGCCCC GGTCTGGCTC 2340  
 CAGGTGACGA CGTGGGTGTA GACGCCCTCT AGCGCCAGGG AGGGGGCGAG GCGCGGGCGT 2400  
 ATGCGGTTGG CCGAAAGTAC GCGCGCACG GACGCCTCGA GGGCCCGCGG GCGCTCCTGG 2460  
 ATCGCGCCGT GCGCGGCGTC CGCGTCCCCG GGGTCCACGT TGAACAGCCC CCAGAACGCA 2520  
 GCCCCGGTGC CGCCGCAGAC CGCAAATTC ACCGAGCTGG CCGTCTGCTC GATCTGCAGG 2580  
 20 CAGACGGCGG CCATGACCCC GCCGAGCAGC TGCCGGAGCG CGGGGCAGGC GTCGCACGCG 2640  
 TCCGGCACCA GCGGCTCCAG CACGGCCCCG GCCCAGGGCT CCGAGGGGGG GGCCGCCACC 2700  
 AGCGCGTCCA GCCTTTCCAG GCCCGCCCCG CCCCAGGCTT CCGGCAGCCC GGCTCCCCG 2760  
 AGGCCCCGGA GGGCGGCCAG GAGCTGGGCC TGGAGCCCGG AGAAACAAAA CCGCGCCGTC 2820  
 CAGACCGGCC CGACGGCCGC CGGGGGGTCG AGTAGTTGGA TGGTGGTGGC CGTGGGGTGC 2880  
 25 CACCGCGCCA CCGCTTCCCG AAAGGCGGGC AGGAGCGGC CGGCCGCCTC CGAGGCCACG 2940  
 GCCGCCATG CCCGCGGGG CAGGACGACC CTGGCGCCA CCGCGGGCCA GGCCCCCAGG 3000  
 CACG 3005

## (2) INFORMATION FOR SEQ ID NO:169:

30

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 221 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

35

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

40

Xaa Xaa Xaa Xaa Xaa Ala Leu Glu Arg Glu Gln Arg Ala Ala Asp Arg  
 1 5 10 15  
 Ala Ala Gly Gly Gly Ala Gly Arg Pro Ala Glu Ala Asp Leu Leu Arg

```

                20                25                30
Ala Asp Tyr Asp Ile Ile Asp Val Ser Lys Ser Met Asp Asp Asp Thr
                35                40                45
Tyr Val Ala Asn Ser Phe Gln His Gln Tyr Ile Pro Ala Tyr Gly Gln
5      50      55      60
Asp Leu Glu Arg Leu Ser Arg Leu Trp Glu His Glu Leu Val Arg Cys
65      70      75      80
Phe Lys Ile Leu Arg His Arg Asn Asn Gln Gly Gln Glu Thr Ser Ile
                85      90      95
10     Ser Tyr Ser Ser Gly Ala Ile Ala Ser Phe Val Ala Pro Tyr Phe Glu
                100     105     110
Tyr Val Leu Arg Ala Pro Arg Ala Gly Ala Leu Ile Thr Gly Ser Asp
                115     120     125
Val Ile Leu Gly Glu Glu Glu Leu Trp Glu Ala Val Phe Lys Lys Thr
15     130     135     140
Arg Leu Gln Thr Tyr Leu Thr Asp Val Ala Ala Leu Phe Val Ala Asp
145     150     155     160
Val Gln His Ala Ala Leu Pro Arg Pro Pro Ser Pro Thr Pro Ala Asp
                165     170     175
20     Phe Arg Ala Ser Asp Arg Gly Gly Ser Arg Ser Arg Thr Arg Thr Arg
                180     185     190
Ser Arg Ser Pro Gly Arg Thr Pro Arg Gly Ala Pro Asp Gln Gly Trp
                195     200     205
Gly Val Glu Arg Arg Asp Gly Arg Pro His Ala Arg Arg
25     210     215     220

```

## (2) INFORMATION FOR SEQ ID NO:170:

## (i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 302 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 35 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

```

Val Arg Arg Thr Arg Ala Gly Asn Ala Gly Met Ala Asp Pro Thr Pro
40  1      5      10      15
Ala Asp Glu Gly Thr Ala Ala Ala Ile Leu Lys Gln Ala Ile Ala Gly
                20      25      30
Asp Arg Ser Leu Val Glu Val Ala Glu Gly Ile Ser Asn Gln Ala Leu
                462

```

```

          35          40          45
Leu Arg Met Ala Cys Glu Val Arg Gln Val Ser Asp Arg Gln Pro Arg
  50          55  ..          60
Phe Thr Ala Thr Ser Val Leu Arg Val Asp Val Thr Pro Arg Gly Arg
5  65          70          75          80
Leu Arg Phe Val Leu Asp Gly Ser Ser Asp Asp Ala Tyr Val Ala Ser
          85          90          95
Glu Asp Tyr Phe Lys Arg Cys Gly Asp Gln Pro Tyr Gly Phe Ala Val
          100          105          110
10 Val Val Leu Thr Ala Asn Glu Asp His Val His Ser Leu Ala Val Pro
          115          120          125
Pro Leu Val Leu Leu His Arg Leu Ser Leu Phe Arg Pro Thr Asp Leu
          130          135          140
Arg Asp Phe Glu Leu Val Cys Leu Leu Met Tyr Leu Glu Asn Cys Pro
15 145          150          155          160
Arg Ser His Ala Thr Pro Ser Leu Phe Val Lys Val Ser Ala Trp Leu
          165          170          175
Gly Val Val Ala Arg His Asp Phe Glu Arg Val Arg Cys Leu Leu Leu
          180          185          190
20 Arg Ser Cys His Trp Ile Leu Asn Thr Leu Met Cys Met Ala Gly Val
          195          200          205
Lys Pro Phe Asp Asp Glu Leu Val Leu Pro His Trp Tyr Met Ala His
          210          215          220
Tyr Leu Leu Ala Asn Asn Pro Pro Pro Val Leu Ser Ala Leu Phe Cys
25 225          230          235          240
Ala Thr Pro Gln Ser Ser Ala Leu Gln Leu Pro Gly Pro Val Pro Arg
          245          250          255
Thr Asp Cys Val Ala Tyr Asn Pro Ala Gly Val Met Gly Ser Cys Trp
          260          265          270
30 Lys Ser Lys Asp Leu Arg Ser Ala Leu Val Tyr Trp Trp Leu Ser Gly
          275          280          285
Ser Pro Lys Arg Arg Thr Ser Ser Leu Phe Tyr Arg Phe Cys
          290          295          300

```

35 (2) INFORMATION FOR SEQ ID NO:171:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 402 amino acids
- (B) TYPE: amino acid
- 40 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

```

5   Ala Cys Leu Gly Ala Trp Pro Ala Val Gly Ala Arg Val Val Leu Pro
    1           5           10           15
    Pro Arg Ala Trp Pro Ala Val Ala Ser Glu Ala Ala Gly Arg Leu Leu
      20           25           30
    Pro Ala Phe Arg Glu Ala Val Ala Arg Trp His Pro Thr Ala Thr Thr
      35           40           45
10  Ile Gln Leu Leu Asp Pro Pro Ala Ala Val Gly Pro Val Trp Thr Ala
    50           55           60
    Arg Phe Cys Phe Ser Gly Leu Gln Ala Gln Leu Leu Ala Ala Gly Leu
    65           70           75           80
    Gly Glu Ala Gly Leu Pro Glu Arg Arg Ala Gly Leu Glu Arg Leu Asp
15           85           90           95
    Ala Leu Val Ala Ala Ala Pro Ser Glu Pro Trp Ala Arg Ala Val Leu
      100           105           110
    Glu Arg Leu Val Pro Asp Ala Cys Asp Ala Cys Pro Ala Leu Arg Gln
      115           120           125
20  Leu Leu Gly Gly Val Met Ala Ala Val Cys Leu Gln Ile Glu Gln Thr
    130           135           140
    Ala Ser Ser Val Lys Phe Ala Val Cys Gly Gly Thr Gly Ala Ala Phe
    145           150           155           160
    Trp Gly Leu Phe Asn Val Asp Pro Gly Asp Ala Asp Ala Ala His Gly
25           165           170           175
    Ala Ile Gln Asp Ala Arg Arg Ala Leu Glu Ala Ser Val Arg Ala Val
      180           185           190
    Leu Ser Ala Asn Gly Ile Arg Pro Arg Leu Ala Pro Ser Leu Ala Leu
      195           200           205
30  Glu Gly Val Tyr Thr His Val Val Thr Trp Ser Gln Thr Gly Ala Trp
    210           215           220
    Phe Trp Asn Ser Arg Asp Asp Thr Asp Phe Leu Gln Gly Phe Pro Leu
    225           230           235           240
    Arg Gly Pro Ala Tyr Ala Ala Ala Ala Glu Val Met Arg Asp Ala Leu
35           245           250           255
    Arg Arg Ile Leu Arg Arg Pro Ala Ala Gly Pro Pro Glu Glu Ala Val
      260           265           270
    Cys Ala Arg Ile Met Glu Asp Ala Cys Asp Arg Phe Val Leu Asp Ala
      275           280           285
40  Phe Gly Arg Arg Leu Asp Ala Glu Tyr Trp Ser Val Leu Thr Pro Pro
    290           295           300
    Gly Glu Ala Asp Asp Pro Leu Pro Gln Thr Ala Phe Arg Gly Gly Ala
    305           310           315           320

```

[illegible]

(2) INFORMATION FOR SEQ ID NO:172:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 428 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

25	CGCGACGCGG	GCCGCTGGGT	CCGCGGACCG	GAGAACGACG	TCCGCGGTCC	GCGGGGCGTA	60
	CCCGGACCCC	ATGGCCAGCC	TGTCGCCGCG	ACCCCGGCGG	CCCCGCCGAC	ACCACCACCA	120
	CCACCACCGC	CGCCGCCGCC	GGCGCGCCCC	CCGCCGGCGC	TCGACCGCCT	CTGACTCATC	180
	AAAATCCGGA	TCCTCGTCTG	CGGCGTCCTC	CGCCTCCTCC	TCCGCCTCCT	CCTCCTCGTC	240
	TGCATCCGCC	TCCTCGTCTG	ACGACGACGA	CGACGACGCC	GCCCGCGCCC	CCGCCAGCGC	300
30	CGCAGACCAC	GCCGCGGGCG	GGACCTCGG	CGCGGACGAC	GAGGAGGCGG	GGGTGCCCCG	360
	GAGGGCCCCG	GGGGCGGCGC	CCCGGCCGAG	CCCGCCAGG	GCCGAGCCCG	CCCCCGGGGC	420
	CGGGGCG						428

(2) INFORMATION FOR SEQ ID NO:173:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15900 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

465



	CGCGCTCCGT	GTGGACGATC	GCCCCGTCGC	CTGGCTGATA	TAGTCCTCGG	GGCGCGCGGG	60
	GCGGGGGGAA	AGGAGGAGGA	CGCGGAGGAG	GAGCGATCGA	CGCCGCCGCG	CCCCGGCTCG	120
	CCGGGGTTCC	GCCCCCAGGT	GGAACCGCAT	TATGCGCGGC	CCCGCCCCGA	CGCCCGCGCG	180
5	TCCGCGTCCG	TGGCGGCGGC	CCGTTGGTCG	CGCCGCCGCC	GGCTCCGCCC	GCGCGGCATC	240
	TCATTAGCGC	CCGGCGCGGG	CGGCTTCCGC	TTCGCCCCGC	GATGCTAATG	AGACCTTCGT	300
	CGCGGGCGGG	CTCGCTCCCC	TGCCCCTCCG	GGTTCGTGGT	AATGAGATGC	CGGCCCCGCG	360
	CTCCCGTTGG	CCCCCGCCGG	CCCCTTTGGG	GCCGGCGAGG	TCGCCCCGTT	GGTCCGCGGG	420
	CGGCTCCGCC	CCAAAGGGGG	CGGGGCCGCA	GGGTAAAAGA	AGTGAGAACG	CGAAGCGTTC	480
10	GCACTTCGTC	CTAATAGTAT	ATATATTATT	AGGGCAAAGT	GCGAGCACTG	GCGCCCTGCC	540
	CGGGGCCCGC	GTCATCCCGC	GGGCTCCGCC	CCAAAGGGGG	CGGGGCCGCA	GGGTAAAAGA	600
	AGTGAGAACG	CGAAGCGTTC	GCACTTCGTC	CTAATAGTAT	ATATATTATT	AGGGCAAAGT	660
	GCGAGCACTG	GCGCCCTGCC	CGGGGCCCGC	GTCATCCCGC	GGGCCCCGCC	CCGAGGCGGG	720
	CCCGGACGGG	GGGCGGGCCG	TTCTCGCGC	ACATAAAGGG	CCGGCGTCCC	GGTCGCGCGC	780
15	GCACCAGGGG	CACACCGGCT	GCGCGGCGGA	GACCGGGACG	GCAGCGGCGG	CATCGCGAAG	840
	GGGGCCACAG	CGAGACAGAG	ACGCCGGCGG	CGAGCGGGGG	ACCGACGCAC	CCGGATCGGA	900
	TCGATACAG	AGACGCGGGC	GCATCGGTTT	CTTTTCGTTT	TGCCTTTCCC	TCCCCCCCCC	960
	CCCCACCTG	TACGTACCGC	GAGGACCCAT	CCACCCACTG	CAGCCTTATC	GCAGGTACGG	1020
	TGACCCGGGG	GCGCCGGGGC	GGGGGGACGG	GACGGGGGGA	CGGGACGGGG	GGACGGGACG	1080
20	GGGGGACGGG	ACGGGGGGAC	GGGACGGGGG	GACGGGACGG	GGGGACGGGA	CGGGGGGACG	1140
	GGACGGGGGG	ACGGGACGGG	GGGACGGGAC	GGGGGGACGG	GACGGGGGGA	CGGGACGGGG	1200
	GGACGGGACG	GGGGGACGGG	ACGGGGGGAC	GGGACGGGGG	GACGGGACGG	GGGGACGGGA	1260
	CGGGGGGACG	GGACGGGGGG	ACGGGACGGG	GGGACGGGAC	GGGGGGACGG	GACGGGGGGA	1320
	CGGGACGGGG	GGACGGGACG	GGGGGACGGG	ACGGGGGGAC	GGGACGGGGG	GACGGGACGG	1380
25	GGGGACGGGA	CGGGGGGACG	GGACGGGGGG	ACGGGACGGG	GGGACGGGAC	GGGGGGGCCC	1440
	CGATCCCAAC	ATCCGCGCTT	TCTCGCAGGC	CGGGCGCCGC	CTTCGTGGAC	GGGACACCGG	1500
	TGTGGTAAC	GGCGACAAGG	CGTCGCCACT	ATGGCAGACA	TCCCCCGGGA	CCCGCCCGCG	1560
	CTCAACACGA	CGCCTGCGAA	TCATGCTCCC	CCATCCCCAC	CCCCGGGTTC	ACGGAAGCGC	1620
	AGACGCCCCG	TCCTCCCCAG	CTCGTCGGAA	TCTGAGGGTA	AGCCCGACAC	AGAATCGGAA	1680
30	TCCTCCTCGA	CCGAGTCGTC	CGAGGATGAG	GCGGGAGACC	TACGCGGCGG	GCGCCGTCGC	1740
	TCCCCGCGGG	AGCTCGGGGG	GAGGTATTTT	TTGGATCTGT	CGGCAGAATC	GACCACGGGG	1800
	ACGGAATCGG	AGGGAACGGG	GCCGTCGGAC	GACGATGATG	ATGATGCGTC	AGACGGCTGG	1860
	TTGGTTGACA	CCCCCCCCCG	TAAATCCAAG	CGACCCCGAA	TCAACCTGCG	ATTAACGAGC	1920
	TCCCCCGACC	GGCGCGCGGG	TGTGGTTTTC	CCCAGAGTGT	GGAGAAACGA	CAGACCTATC	1980
35	CGCGCGGCGC	AACCCAGGC	CCCGGCCAG	TCTTCGGGG	ATCGCGCAGC	CGCACC GCGG	2040
	CGCTCTGCTC	GCCAGGCCCA	GATGCGGAGC	GGAGCCGCCT	GGACGCTTGA	TCTGCATTAC	2100
	ATACGCCAGT	GCGTCAACCA	GCTCTTTTCG	ATCCTGCGTG	CCGCCCCGAA	CCCGCCCGGC	2160
	AGCGCCAACC	GCCTGCGCCA	CCTGGTGCGA	GA CTGCTACC	TTATGGGCTA	CTGCCGACC	2220
	CGCTGGGGG	CGCGCACGTG	GGGCCGCTTG	CTGCAGATCT	CGGGCGGAAC	CTGGGACGTG	2280
40	CGCTGCGAA	ACGCAATCCG	GGAGGTCGAG	GCGCGTTTTG	AACCCGCCGC	CGAGCCCGTG	2340
	TGCGAGCTGC	CCTGTCTGAA	CGCCAGGCGT	TACGGCCCCG	AGTGTGATGT	TGGCAATCTC	2400
	GAGACCAACG	GCGGCTCGAC	GAGCGATGAT	GAGATATCGG	ATGCGACGGA	CTCGGACGAT	2460
	ACCCTCGCGT	CCCATTCGGA	CACGGAGGGG	GGGCCCTCCC	CGGCCGCGCG	GGAGAACCCG	2520

	GAATCCGCGT	CCGGCGGGGC	TATCGCGGCT	CGGC TGGAGT	GTGAGTTTGG	GACGTTTGAC	2580
	TGGACGTCCG	AGGAGGGCTC	CCAGCCCTGG	CTGTCCGCGG	TGGTCGCCGA	TACCAGCTCC	2640
	GCCGAACGCT	CTGGCCTACC	CGCCCGGGGC	GCGTGTGCGG	CAACGGAAGC	CCCAGAACGC	2700
	GAGGACGGGT	GCCGAAAAAT	GCGCTTCCCC	GCCGCCTGCC	CCTATCCCTG	CGGCCACACA	2760
5	TTTCTCCGGC	CATGAGCGCG	GGACCCCCAG	CCCGGTGTGT	TTGCCAAACG	AAAATAAACG	2820
	CCCTACAAGA	AAGCTTTTGT	GTCTGAGTGT	CTGGTTTTC	TGGGGGTGGA	GGAAGGAACG	2880
	ACAAAAAAG	AAACAAACGC	GACACCGCTC	GTACGTGTAA	TGGGGCGCAG	TGTTTTTTAT	2940
	TAGCATCGGG	GGGGGGTTAG	AGGTTGGTGA	TTGGATAGCA	AACGTGGGAT	GACGGAGGCC	3000
	ACTCGTCGCC	AACGGCCAGC	GGGGGCCCGG	GGTTC TGGGG	GTCATCGTCC	CCCGTCTGCC	3060
10	AGGAGGGCTC	ATCGGGAATC	TCGGGTCGCC	CCATGCACGT	AAAACACGGG	CGCTGCGTGG	3120
	GGTGGGTGCG	CGGATGCGGG	CGGGATGATG	CGGGGCGGGG	TTTGTGTGA	GGAGCCACGA	3180
	GGGACCGTAG	CCAGCGAAGA	CAGCTGCGTT	CCCGGTCGCC	GGGCACCACC	ACGCCGTATT	3240
	GGTATTCGTA	TCGGCTAAGG	AGATTTTCCA	GGGGGTGATT	AGGCGCTGCG	GGGAACGGGG	3300
	TCCACGACAC	GGTCCGCTCG	GGCAAAAACC	GATCGGGCAG	GGGCCACGGT	TCCCCACCC	3360
15	ACGCGTCGTT	GGTCTTCATG	GCGATGAAGC	GAAACCCAG	CCGGGTTTTT	TGTGCGTACT	3420
	CTAAAAACGG	CACACACAGG	TCCGCCGCC	CGACCACCCA	CAGGTGGTAT	AGCCGGTGGG	3480
	GGCCGGGGCG	CTCTTGATGC	AGGAGCCGAA	AACACGCAGG	GGCATCCAGA	ATCTCGATGC	3540
	TTTCCAGGGG	GTCGTCCTCC	GCAAACAGGC	CCGTCGTGGT	GTTTGGGGGA	CAGCGACAGG	3600
	AGCGGGTTCG	CACGATCGGT	CGGGTGAATT	TGGGCAAGTC	CATCAGAGGC	TCGGCCAGCC	3660
20	TGCGAAGGTT	CGCCGGGCGA	ACCACCACCG	GGGTTCCCAG	AGGCTCGGAG	GCCAGGATCC	3720
	GGCATTGCCG	AAGCAGAAAA	CTCCACAGAG	CCGGGCTTGC	GTCAGCGGAA	GTCGCGGGCA	3780
	GGGCGTTTCG	TTGGTCTAGG	AGGGTAACCA	CAC TTACAAC	AACAACGCC	ATGTCGGTAT	3840
	ATTAGGCCCG	TGGTCCGATC	TTCACTCACT	CGCTGTCTG	CGGACCTATG	CACGGCGGGA	3900
	CGGCGCGCGG	ACCCGGGGGG	GCTGCTTGCT	ATCACACGGC	CCGTTCGCAC	GTTTCGATTTT	3960
25	TTCAGCCTTG	TTTGGTTGGC	TAGGTATCCC	GGATAATCTG	ACGTTCCGGA	TATAGGGGGC	4020
	GGGGGTAGTG	GGGGGTGTG	TCGACAAACT	GCCGCTTCTT	AAAACACCGG	GGCCCGTTCG	4080
	TCGGGGTGCT	CGTTGGTTGG	CACGCGCGAC	GCGGCGAATG	GCCTGTCGTA	AGTTCTGTGG	4140
	GGTCTACCGT	AGACCCGACA	AGAGACAGGA	GGCGTCCGTC	CCGCCGGAGA	CAAACACGGC	4200
	CCCGGCCCTC	CCGCGGAGCA	CCTTTTATAC	CCCCGCGGAG	GATGCGTACC	TGGCCCCCGG	4260
30	GCCCCCGGAA	ACCATCCACC	CTTCCCGCCC	ACCGTCCCCC	GGCGAGGCTG	CGCGCCTGTG	4320
	TCAGCTGCAG	GAGATCTTGG	CCCAGATGCA	CAGCGACGAG	GACTACCCCA	TCGTGGACGC	4380
	CGCGGGTGCG	GAGGAGGAAG	ACGAGGCCGA	CGATGACGCC	CCGGATGACG	TGGCCTACCC	4440
	GGAGGACTAC	GCGGAGGGGC	GT TTTCTGTC	CATGGTTTCG	GCCGCCCCC	TGCCCCGAGC	4500
	CAGCGGCCAT	CCTCCTGTTC	CGGGCCGCGC	AGCCCCCCCC	GACGTCCGGA	CCTGCGACAG	4560
35	CGGTAAGGTG	GGGGCCACGG	GGTTCACCCC	GGAAGAGCTC	GACACCATGG	ACCGGGAGGC	4620
	ACTTCGGGCC	ATCAGCCGCG	GGTGCAAGCC	CCCTTCGACC	CTGGCAAAAC	TGGTGACCGG	4680
	GCTGGGATTC	GCGATCCACG	GAGCGCTCAT	CCCGGGGTGCG	GAGGGGTGTG	TCTTTGATAG	4740
	CAGCCACCCG	AACTACCCTC	ATCGGGTAAT	CGTCAAGGCG	GGGTGGTACG	CCAGCACGAA	4800
	CCACGAGGCG	CGGCTGCTGA	GACGCCTGAA	CCACCCCGCG	ATCCTACCCC	TCCTGGACCT	4860
40	GCACGTCGTT	TCTGGGGTCA	CGTGTCTGGT	CCTCCCCAAG	TATCACTGCG	ACCTGTATAC	4920
	CTATCTGAGC	AAGCGCCCGT	CTCCGTTGGG	CCACCTACAG	ATAACCGCGG	TCTCCCGGCA	4980
	GCTCTTGAGC	GCCATCGACT	ACGTCCACTG	CGAAGGCATC	ATCCACCGCG	ATATTAAGAC	5040
	CGAGAACATC	CTCATCAACA	CCCCCGAGAA	CATCTGTCTG	GGGACTTTG	GGGCGGCGTG	5100

	CTTTGTGCGC	GGGTGTCGAT	CGAGCCCCTT	CCATTACGGG	ATCGCAGGCA	CCATCGATAC	5160
	AAACGCCCCC	GAGGTCTTGG	CCGGGGATCC	GTACACCCAG	GTAATCGACA	TCTGGAGCGC	5220
	CGGCCTGGTG	ATCTTTGAGA	CCGCCGTCCA	CACCGCGTCC	TTGTTCTCGG	CCCCGCGCGA	5280
	CCCCGAAAGG	CGGCCGTGCG	ACAACCAGAT	CGCGCGCATC	ATCCGACAGG	CCCAGGTACA	5340
5	CGTCGACGAG	TTTCCAACGC	ACGCGGAATC	GCGCCTCACC	GCGCACTACC	GCTCGCGGGC	5400
	GGCCGGGAAC	AATCGTCCGG	CGTGGACCCG	ACCGGCATGG	ACCCGCTACT	ACAAGATCCA	5460
	CACAGACGTC	GAATATCTCA	TCTGCAAAGC	CCTTACCTTT	GACGCGGCGC	TCCGCCCAAG	5520
	CGCCGCGGAG	TTGCTGCGCC	TGCCGTATT	TCACCCTAAG	TGACCCCGCT	CCCCCGGGG	5580
	GGCGTGGAGG	GGGGGCTGGT	TGGATGTTTT	TGCACAAAAA	GACGCGGCCC	TCGGGCTTTG	5640
10	GTGTTTTTGG	CACCTTGCCG	CCCGGCGTCA	TGCACGCCAT	CGCTCCCAGG	TTGCTTCTTC	5700
	TTTTTGTTCT	TTCTGGTCTT	CCGGGGACAC	GCGGCGGGTC	GGGTGTCCCC	GGACCAATTA	5760
	ATCCCCCAA	CAACGATGTT	GTTTTCCCGG	GAGGTTCCCC	CGTGGCTCAA	TATTGTTATG	5820
	CCTATCCCCG	GTGGGACGAT	CCCGGGCCCT	TGGGTTCGCG	GGACGCCGGG	CGGCAAGACC	5880
	TGCCCCGGCG	CGTCGTCCGT	CACGAGCCCC	TGGGCCGCTC	GTTCCTCACG	GGGGGGCTGG	5940
15	TTTTGCTGGC	GCCGCCGGTA	CGCGGATTTG	GCGCACCCAA	CGCAACGTAT	GCGGCCCGTG	6000
	TGACGTACTA	CCGGCTCACC	CGCGCCTGCC	GTCAGCCCAT	CCTCCTTCGG	CAGTATGGAG	6060
	GGTGTCGCGG	CGGCGAGCCG	CCGTCCCCAA	AGACGTGCGG	GTCTACACG	TACACGTACC	6120
	AGGGCGGCGG	GCCTCCGACC	CGGTACGCTC	TCGTAAATGC	TTCCCTGCTG	GTGCCGATCT	6180
	GGGACCGCGC	CGCGGAGACA	TTGAGTACC	AGATCGAACT	CGGCGGCGAG	CTGCACGTGG	6240
20	GTCTGTTGTG	GGTAGAGGTG	GGCGGGGAGG	GCCCCGCCCC	CACCGCCCCC	CCACAGGCGG	6300
	CGCGTGCGGA	GGGCGGCCCC	TGCGTCCCCC	CGGTCCCCGC	GGGCCGCCCC	TGGCGCTCGG	6360
	TGCCCCCGGT	ATGGTATTCC	GCCCCCAACC	CCGGGTTCG	TGGCCTGCGT	TTCCGGGAGC	6420
	GCTGTCTGCC	CCCACAGACG	CCCGCCGCCC	CCAGCGACCT	ACCACGCGTC	GCTTTTGCTC	6480
	CCCAGAGCCT	GCTGGTGGGG	ATTACGGGCC	GCACGTTTAT	TCGGATGGCA	CGACCCACGG	6540
25	AAGACGTCGG	GGTCCTGCCA	CCCCATTGGG	CCCCCGGGGC	CCTAGATGAC	GGTCCGTACG	6600
	CCCCCTTCCC	ACCCCGCCCC	CGGTTTCGAC	GCGCCCTGCG	GACAGACCCC	GAGGGGGTCG	6660
	ACCCCGACGT	TCGGGCCCCC	CTAACCGGGC	GGCGCCTCAT	GGCCTTGACC	GAGGACGCGT	6720
	CCTCCGATTG	GCCTACGTCC	GCTCCGGAGA	AGACGCCCCC	CCCTGTGTCG	GCCACCGCCA	6780
	TGGCGCCCTC	AGTCGACCCA	AGCGCGGAAC	CGACCGCCCC	CGCAACCACT	ACTCCCCCGG	6840
30	ACGAGATGGC	CACACAAGCC	GCAACGGTCG	CCGTTACGCC	GGAGGAAACG	GCAGTCGCCT	6900
	CCCCGCCCCG	GA CTGCATCC	GTGGAGTCGT	CGCCACTCCC	CGCCGCGGCG	GCAACGCCCC	6960
	GGGCGGGGCA	CACGAACACC	AGCAGCGCCC	CCGACGCGAA	AACGCCCCCC	ACCACACCAG	7020
	CCCCCACGAC	CCCCCGCCCC	ACGTCTACCC	ACGCGACCCC	CCGCCCCACG	AGTCCGGGGC	7080
	CCCAAACAAC	CCCTCCCGGA	CCCGCAACCC	CGGGTCCGGT	GGGCGCCTCC	GCCGCACCCA	7140
35	CGGCGGATTG	CCCCCTCACC	GCCTCGCCCC	CCGCTACCGC	GCCGGGGCCC	TCGGCCGCCA	7200
	ACGTTTTCGGT	CGCCGCGACC	ACCGCCACGC	CCGGAACCCG	GGGCACCGCC	CGTACCCCCC	7260
	CAACGGACCC	AAAGACGCAC	CCACACGGAC	CCGCGGACGC	TCCCCCGGGC	TCGCCAGCCC	7320
	CCCCACCCCC	CGAACATCGC	GGCGGACCCG	AGGAGTTTGA	GGGCGCCGGG	GACGGCGAAC	7380
	CCCCCGATGA	CGACGACAGC	GCCACCGGCC	TCGCTTCCG	AACTCCGAAC	CCCAACAAC	7440
40	CACCCCCCGC	GCGCCCCGGG	CCCATCCGCC	CCACGCTCCC	GCCAGGAATT	CTTGGGCCGC	7500
	TCGCCCCCAA	CACGCCTCGC	CCCCCGCCCC	AAGCTCCCGC	TAAGGACATG	CCCTCGGGCC	7560
	CCACACCCCA	ACACATCCCC	CTGTTCTGGT	TCCTAACGGC	CTCCCTTGCT	CTAGATATCC	7620
	TCTTTATCAT	CAGCACCACC	ATCCACACGG	CGGCGTTCGT	TTGTCTGGTC	GCCTTGCCAG	7680

	CACAACTTTG	GCGCGGCCGG	GCGGGGCGCA	GGCGATACGC	GCACCCGAGC	GTGCGTTACG	7740
	TATGTCTGCC	ACCCGAGCGG	GATTAGGGGG	TGGGGTGGGG	GCGAGAAACG	ATGAAGGACG	7800
	GGAAAGGGAA	CAGCGACCAA	ATGCCACGAT	AAGAACAATA	AACCTGTGAC	GTCAATCGGA	7860
	TATGTGAGTT	TGGTTGTGTT	TTGTGGGACT	GGGGGCGGGG	GGTGGGAGGT	ATCAGTGGGT	7920
5	GACAGAGTCT	TTTAAAAGAC	GTGTCCCGGG	GCCCTCGAGA	CGCGCAACTT	TTGGCCACAC	7980
	AGAGAAAGGC	CCCCAGACGA	AGTCACCCGG	GTCCCCGAAC	AAAAACAAAA	ACCTTGACCG	8040
	CCGCCGGGGG	GCGTGCCGTG	TGTTTTGGTC	TCAATGGATC	GGTATGCCGT	TCGGACCTGG	8100
	GGGATTGTGG	GAATCCTCGG	GTGTGCTGCT	GTTGGGGCCG	CACCCACCGG	CCCCGCGTCC	8160
	GATACAACAA	ACGCGACCGC	ACGCCTCCCC	ACGCACCCCC	CACTCATCCG	TTCCGGGGGC	8220
10	TTTGCCGTCC	CCCTCATCGT	GGGGGGGCTG	TGTCTCATGA	TTCTGGGGAT	GGCGTGTCTA	8280
	CTCGAGGTCC	TGCGTCGCCT	GGGTCGCGAG	TTGGCGAGGT	GCTGCCCCCA	CGCGGGCCAA	8340
	TTTGCCCCAT	GATTTTTCGC	CTTCTGCGC	TTGCCCCAC	CCCATCGCCC	CGATTGTGTG	8400
	TCGGGTGCCC	GGGTACAGC	AGCTATGGAG	CGGTCGGTAA	TATAACTTTG	GTTGTCGCCA	8460
	CACGCCCCGT	GCCGGGCATG	GGTTGTGCGG	GAAAGACGAA	ATAATCCGGC	GATCCCCAAG	8520
15	CGTACCAACT	TGGGGGGGGG	GGGAAAGAAA	CTAAAAACAC	ATCAAGCCCA	CAACCCATCC	8580
	CACAAGGGGG	GTTATGGCGG	ACCCACCGCA	CCACCATACT	CCGATTTCGAC	CACATATGCA	8640
	ACCAATCAC	CCCCAGAGGG	GAGGTTCAT	TTTTACGAGG	AGGAGGAGTA	TAATAGAGTC	8700
	TTTGTGTTTA	AAACCCGGGG	TCGGTGTGGT	GTTCCGGTCAT	AAGCTGCATT	GCGAACGACT	8760
	AGTCGCCGTT	TTTCGTGTGC	ATCGCGTATC	ACGGCATGGG	GCGTTTGACC	TCCGGCGTCG	8820
20	GGACGGCGGC	CCTGCTAGTT	GTCGCGGTGG	GACTCCGCGT	CGTCTGCGCC	AAATACGCCCT	8880
	TAGCAGACCC	CTCGCTTAAG	ATGGCCGATC	CCAATCGATT	TCGCGGAAG	AACCTTCCGG	8940
	TTTTGGACCA	GCTGACCGAC	CCCCCGGGG	TGAAGCGTGT	TTACCACATT	CAGCCGAGCC	9000
	TGGAGGACCC	GTTCCAGCCC	CCCAGCATCC	CGATCACTGT	GTACTACGCA	GTGCTGGAAC	9060
	GTGCCTGCCG	CAGCGTGCTC	CTACATGCCC	CATCGGAGGC	CCCCCAGATC	GTGCGCGGGG	9120
25	CTTCGGACGA	GGCCCGAAAG	CACACGTACA	ACCTGACCAT	CGCCTGGTAT	CGCATGGGAG	9180
	ACAATTGCGC	TATCCCCATC	ACGGTTATGG	AATACACCGA	GTGCCCCTAC	AACAAGTCGT	9240
	TGGGGGTCTG	CCCCATCCGA	ACGCAGCCCC	GCTGGAGCTA	CTATGACAGC	TTTAGCGCCG	9300
	TCAGCGAGGA	TAACCTGGGA	TTCTGATGC	ACGCCCCCGC	CTTCGAGACC	GCGGGTACGT	9360
	ACCTGCGGCT	AGTGAAGATA	AACGACTGGA	CGGAGATCAC	ACAATTTATC	CTGGAGCACC	9420
30	GGGCCCCGCG	CTCCTGCAAG	TACGCTCTCC	CCCTGCGCAT	CCCCCGGCA	GCGTGCCCTCA	9480
	CTTCGAAGGC	CTACCAACAG	GGCGTGACGG	TCGACAGCAT	CGGGATGCTC	CCCCGCTTTA	9540
	TCCCCGAAAA	CCAGCGCACC	GTCGCCCTAT	ACAGCTTAAA	AATCGCCGGG	TGGCACGGCC	9600
	CCAAGCCCCC	GTACACCAGC	ACCCTGCTGC	CGCCGGAGCT	GTCCGACACC	ACCAACGCCA	9660
	CGCAACCCGA	ACTCGTTCCG	GAAGACCCCG	AGGACTCGGC	CCTCTTAGAG	GATCCCGCCG	9720
35	GGACGGTGTC	TTGCGAGATC	CCCCCAAAC	GGCACATCCC	GTCGATCCAG	GACGTCGCGC	9780
	CGCACCACGC	CCCCGCCGCC	CCCAGCAACC	CGGGCCTGAT	CATCGGCGCG	CTGGCCGGCA	9840
	GTACCCTGGC	GGTGCTGGTC	ATCGGCGGTA	TTGCGTPTTG	GGTACGCCGC	CGCGCTCAGA	9900
	TGGCCCCCAA	GCGCCTACGT	CTCCCCACA	TCCGGGATGA	CGACGCGCCC	CCCTCGCACC	9960
	AGCCATTGTT	TTACTAGAGG	AGTATCCCCG	CTCCCGTGTA	CCTCTGGGCC	CGTGTGGGAG	10020
40	GGTGCTGGG	GTATTTGGGT	GGGACTTGGA	CTCCGCATAA	AGGGAGTCTC	GAAGGAGGGA	10080
	AACTAGGACA	GTTCATAGGC	CGGGAGCGTG	GGGCGCGCAC	CGCTGTCCCG	ACGATTAGCC	10140
	ACCGCGCCA	CAGTCACCTC	GACCCGTCCG	ATCCCGGTAT	GCCCGGCCGC	TCGCTGCAGG	10200
	GCCTGGCGAT	CCTGGGCCCTG	TGGGTCTGCG	CCACCGGCCT	GGTCGTCCGC	GGCCCCACGG	10260

	TCAGTCTGGT	CTCAGACTCA	CTCGTGGATG	CCGGGGCCGT	GGGGCCCCAG	GGCTTCGTGG	10320
	AAGAGGACCT	GCGTGTTTTC	GGGGAGCTTC	ATTTTGTGGG	GGCCCAGGTC	CCCCACACAA	10380
	ACTACTACGA	CGGCATCATC	GAGCTGTTC	ACTACCCCT	GGGGAACCAC	TGCCCCCGCG	10440
	TTGTACACGT	GGTCACACTG	ACCGCATGCC	CCCCCGCCC	CGCCGTGGCG	TTCACCTTGT	10500
5	GTGCTCGAC	GCACCACGCC	CACAGCCCCG	CCTATCCGAC	CCTGGAGCTG	GGTCTGGCGC	10560
	GGCAGCCGCT	TCTGCGGGT	CGAACGGCAA	CGCGCGACTA	TGCCGGTCTG	TATGTCCTGC	10620
	GCGTATGGGT	CGGCAGCGCG	ACGAACGCCA	GCCTGTTTGT	TTTGGGGGTG	GCGCTCTCTG	10680
	CCAACGGGAC	GTTTGTGTAT	AACGGCTCGG	ACTACGGCTC	CTGCGATCCG	GCGCAGCTTC	10740
	CCTTTTCGGC	CCCGCGCCTG	GGACCCTCGA	GCGTATACAC	CCCCGGAGCC	TCCCGGCCCA	10800
10	CCCCCTCCACG	GACAACGACA	TCCCCGTCCT	CCCCCGAGA	CCCGACCCCC	GCCCCCGGGG	10860
	ACACAGGGAC	GCCCCGCGCC	GCGAGCGGCG	AGAGAGCCCC	GCCCAATTCC	ACGCGATCGG	10920
	CCAGCGAATC	GAGACACAGG	CTAACCGTAG	CCCAGGTAAT	CCAGATCGCC	ATACCGGCGT	10980
	CCATCATCGC	CTTTGTGTTC	CTGGGCAGCT	GTATCTGCTT	CATCCATAGA	TGCCAGCGCC	11040
	GATACAGGCG	CCCCCGCGGC	CAGATTTACA	ACCCCGGGG	CGTTTCCTGC	GCGGTCAACG	11100
15	AGGCGGCCAT	GGCCCGCCTC	GGAGCCGAGC	TGCGATCCCA	CCCAAACACC	CCCCCAAAC	11160
	CCCGACGCCG	TTCGTCGTCG	TCCACGACCA	TGCCTTCCCT	AACGTCGATA	GCTGAGGAAT	11220
	CGGAGCCAGG	TCCAGTCGTG	CTGCTGTCCG	TCAGTCCTCG	GCCCCGAGT	GGCCCCGACG	11280
	CCCCCAAGA	GGTCTAGGTC	CAAGCGGGCC	GTTCGGCAGG	CCCCCCCCAC	CGCCCCCATC	11340
	GTGGTTATTT	CCCCCCCAAT	AAACCGATGT	TATTTGCCTA	TATGCGTGTG	TTGGATCCCT	11400
20	TTGTGATCGT	TCGTCATTCC	CCGGATGGCA	TGGGAGGCGG	GTAATGGATG	GGCGGGGCCC	11460
	GGGGGGGGAG	GAAAAAGAAT	AAAGGGGTA	GTGTCGGAGA	GGCCCGCCGC	GCATTTAAGG	11520
	AGTCGCCGCC	CCGACTCTGT	GTCTTCGGGT	GACTTGGTGC	GCCGCCGTCA	GCTAGTCTCC	11580
	GATCTGCCCC	GACCGACGGC	TCCTGCCACC	CGAACATGGC	TCGCGGGGCC	GGGTGGTGT	11640
	TTTTTGTG	AGTTTGGGTC	GTATCGTGCC	TGGCGGCAGC	ACCCAGAACG	TCCTGGAAAC	11700
25	GGGTAACCTC	GGGCGAGGAC	GTGGTGTTCG	TTCCGGCGCC	CGCGGGGCCG	GAGGAACGCA	11760
	CCCGGGCCCA	CAAACTACTG	TGGGCCGCGG	AACCCCTGGA	TGCCTGCGGT	CCCCTGCGCC	11820
	CGTCGTGGGT	GGCGCTGTGG	CCCCCCCCGAC	GGGTGCTCGA	GACGGTCGTG	GATGCGGCGT	11880
	GCATGCGCGC	CCCGGAACCG	CTCGCCATAG	CATACAGTCC	CCCGTTCCCC	GCGGGCGACG	11940
	AGGGACTGTA	TTCGGAGTTG	GCGTGGCGCG	ATCGCGTAGC	CGTGGTCAAC	GAGAGTCTGG	12000
30	TCATCTACGG	GGCCCTGGAG	ACGGACAGCG	GTCTGTACAC	CCTGTCCGTG	GTCGGCCATA	12060
	GCGACGAGGC	GCGCCAAGTG	GCGTCGGTGG	TTCTGGTCTG	GGAGCCCGCC	CCTGTGCCGA	12120
	CCCCGACCCC	CGACGACTAC	GACGAAGAAG	ACGACGCGGG	CGTGAGCGAA	CGCACGCCGG	12180
	TCAGCGTTCC	CCCCCAACC	CCCCCCCCGTC	GTCCCCCGT	CGCCCCCCG	ACGCACCTTC	12240
	GTGTATATCC	CGAGGTGTCC	CACGTGCGCG	GGGTAACGGT	CCATATGGAG	ACCCCGGAGG	12300
35	CCATTCTGTT	TGCCCCCGGG	GAGACGTTTG	GGACGAACGT	CTCCATCCAC	GCCATTGCCC	12360
	ACGACGACGG	TCCGTACGCC	ATGGACGTCG	TCTGGATGCG	GTTTGACGTG	CCGTCTCTGT	12420
	GCGCCGAGAT	GCGGATCTAC	GAAGCTTGTC	TGTATCACC	GCAGCTTCCA	GAGTGTCTAT	12480
	CTCCGGCCGA	CGCGCCGTGC	GCCGTAAGTT	CCTGGGCGTA	CCGCTTGGCG	GTCCGAGCT	12540
	ACGCCGGCTG	TTCCAGGACT	ACGCCCCCGC	CGCGATGTTT	TGCCGAGGCT	CGCATGGAAC	12600
40	CGGTCCCGGG	GTTGGCGTGG	CTGGCTTCCA	CCGTCAATCT	GGAATTCCAG	CACGCCTCCC	12660
	CCAGCACGCG	CGGCCTCTAC	CTGTGCGTGG	TGTACGTGGA	CGATCATATC	CACGCCTGGG	12720
	GCCACATGAC	CATCAGCACC	GCGGCGCAGT	ACCGGAACGC	GGTGGTGGA	CAGCACCTCC	12780
	CCCAGCGCCA	GCCCCGAGCC	GTCGAGCCCA	CCCCCGCCGA	CGTGAGAGCC	CCCCCTCCCG	12840

	CGCCCTCCGC	GCGCGGCCCG	CTGCGCCTCG	GGGCGGTGCT	GGGGGCGGCC	CTGTTGCTGG	12900
	CCGCCCTCGG	GCTGTCCGCG	TGGGGCGTGC	ATGACCTGCT	GGCGCAGGCG	CTCCTGGCGG	12960
	GCGGTTAAAA	GCCGGGCCTC	GGCGAEGGCG	CCCACCTTACA	TTCGCGTGGC	GGACAGCGAG	13020
	CTGTACGCGG	ACTGGAGTTC	GGACAGCGAG	GGGGAGCGCG	ACGGGTCCCT	GTGGCAGGAC	13080
5	CCTCCGGAGA	GACCCGACTC	TCCCTCCACA	AATGGATCCG	GCTTTGAGAT	CTTATCACCA	13140
	ACGGCTCCGT	CTGTATACCC	CCATAGCGAG	GGGCGTAAAT	CTCGCCGCCC	GCTCACCACC	13200
	TTTGGTTTCG	GAAGCCCGGG	CCGTCGTCAC	TCCCAGGCCT	CCTATTGCTC	CGTCTCTTGG	13260
	TAAGCGTCT	TCCGACGACG	CGGACGTCGG	CGATGAAGTG	ATTGCCATCG	CGGACGCACG	13320
	CGGGGACCCG	CCAGAGACCC	TGCCCCCGCG	CGCGGGCGGC	GCCGCGCCCG	CGTGCCGCAG	13380
10	ACCACCTCGC	GGCGGCTCCC	CCGCGGCCCT	TCCCGTGGCC	CTCCACGCCG	TGGACGCCCC	13440
	CTCCAATTTC	GTCACCTGGC	TCGCCGTGCG	CTGGCTGCGG	GGGGCGGTGG	GTCTCGGGGC	13500
	CGTCTGTGTC	GGGATTGCGT	TTTACGTGAC	GTCAATCGCC	CGAGGCGCAT	AAAGGTCCCG	13560
	CGGCCAGCCC	CGCCGCAGCT	CATAAAATC	GTGAGTCACG	GCAACCGCAC	CTTCGCCTCC	13620
	GGCCCTCCGC	CAGCGCCCTT	CCGCGTCCGC	GATGACCTCC	CGGCCCGCCG	ACCAAGACTC	13680
15	GGTGC GTTCC	AGCGCGTCGG	TGCCGCTTTA	CCCCGCGGCC	TCGCCCGTCC	CGGCAGAAGC	13740
	CTACTACTCG	GAAAGCGAAG	ACGAGGCCGC	CAACGACTTC	CTCGTGCGCA	TGGGCCGCCA	13800
	GCAGTCGGTC	CTAAGGCGCC	GACGGCGGCG	CACGCGGTGC	GTGGGGCTGG	TTATCGCCTG	13860
	TCTCGTCGTG	GCCCTCCTAT	CTGGAGGGTT	CGGGGCACTT	TTGGTGTGGC	TGCTCCGCTA	13920
	AATGACGCCT	CGATGTATGG	CGCCTTCTTC	GCCCCACCCC	CTCGCCGCGA	CCCACGTCCG	13980
20	TATGTTAATT	GCAATAAAGT	GGTTGATTGT	CATTACGGTC	TACTAGGTTG	TCTTTTTTTT	14040
	TTGGGGGGGG	GGGGAAGGAA	ATGCAGAAAA	GGGTAAGAAA	TTCTCGGAAT	TTACCCCCC	14100
	GGGGGGGGCA	AGTGCAGTAC	CCCAGTTCCT	CAGTGTTTGG	GAAATCTATT	GAACCTCTCC	14160
	GGCTCCTCCG	TGTTAGGGAA	GTCTCTTGGG	GAAATCTATT	GACCTCTCGC	CCCCCCCCC	14220
	AGGAGGGGGC	AGTGCAGTAC	CCCAGTTCCT	CCGTGCTGGG	GAAATCTCTC	TGCCGGGTAC	14280
25	GGGCTCCAGA	CGAAGGACCC	ATACATTTC	CCATCCGCAC	CCCACATCTG	GCGTTCTAGA	14340
	GTCACGACGC	ATTTGCCCCC	GTCCCCGCAG	CAACACACAA	AGCGATTTCA	ATTTTTCACGA	14400
	TTTTTATTATT	AATTACACCA	ACCACCCTGT	CCCCGGGACG	TGGTCAGGAC	CGGGGGTCCG	14460
	CACCCAAACG	CACGAAACAA	ATGCTGGCAG	TGTGCCGAAT	ATAACCCCGC	GTAGGAACAC	14520
	GTCGACGCGT	GCGCCAAACA	GCACCAGAAG	GCGCATGCCA	TCAGCAGGTC	GTGCATATGG	14580
30	CGATGTGTTT	GGACGCAGGG	CGCAGCCGCG	GCGATAAAAT	TCATGGCGGC	CGTCCGCCAG	14640
	GGCCACAGCG	GCGAGGACTC	CCTGTTGGCC	CGAAGCCATT	GGGTATGAAC	CAGCTGCGCC	14700
	TCCTGTCCGA	CCCTGGCTCC	CGCCAGCGGG	GGCGGTGGGT	CGTGGGTGTT	GAGAGCACAC	14760
	AGGCGGGACA	CCTCGATCAC	CGTCCGAAAA	AAGGCCCGGT	GGTCCGCGGG	CAGCATCTGC	14820
	AGGTGCGCCA	GGGCCTGGGC	GTTGAGAGGG	TACAACTCGG	AGCCGGGGGA	CTCCGGGGGC	14880
35	CGGTCCGCGC	GGTGCCGCGA	GTGGGCACGC	TTTGGGGCCC	GGGTGTCGGA	CGCGGGCGCG	14940
	TTACGGATCC	CGACGCGGGG	CAGAACGTAC	GTGCGTTGGC	GCGGCGATGA	GGGGTCCGGG	15000
	CTGCCGAGGG	GGGCGTAGGG	GACCGGGCTA	GGCAAGCCCG	CGGGTTGCGC	GGGGTTCCCG	15060
	TGGGGGTCTA	GGCTCCCTGG	GCACCCGTGG	GGGTCGTGGG	GGTCGCGGGT	CCCTGGGTAT	15120
	GCGCGGGACC	CTGGGTTCTC	TGGGAGATCG	TGGAACTCGC	GGTTCCCTGG	GCTCTCGGGG	15180
40	AACCCGGGGC	TCCCTGGGGA	CACGTGGTGC	CCTGGGAATT	CTTGATGGTC	GGACGGCTTC	15240
	AGATGGCTTC	GGGATCGAGA	GGGCCGCACA	GACTCGTAGT	AGACCCGAAT	CTCCACGTTT	15300
	CCCCGCCGCC	GGATCATGGT	CGCCGCCCCG	GTGCGGGGGC	CCGTCGGTCG	GAAGCGAGTG	15360
	CCCTTCAAGC	GTGTCCGCTC	CTCTGGGCTG	CATGCCGTCG	GATGGGGTGC	CTTTTAAGGA	15420

AAGGTCTCGG CTGCCCCGCC CAACCGGGGT TTGGGGGTGG GCCGGGGAAA CCCC GGATGC 15480  
 CATGGGGGGG GTCACACCCT AAGCGCCGGC GCGCTGGTTG GGTGGGGGTA GAGGGGAGTC 15540  
 CCCGGTTCGAC GAGATCGTAT CAAGGGGCCA GCACGCGATC CTGCCGCTCG TTCGATCTAG 15600  
 CACACCCACG GGTCTGCTGT GTGGGATTTC GACTCGCGGG ATCCGATCGC ACGTCCGGAG 15660  
 5 GACACAGCAG CGGGAGCTCC GGGTCGGTCA CCGCAGTTCT GGCCGCCTCT CGGTCTCTCC 15720  
 GTTCCCTTTT ATGGATCTCC GCGCAGACAT CGCCATACGT CCGGTGTGTG CACCGCGAAG 15780  
 AATCCAAAAA CATGTCCGTC GTTTTCAGGG CCAAGACAT GGTGTCCCGT CCACGAAGGC 15840  
 GCGCCCCGGC CTGCGAGAAA GCGCGGATGT TGGGATCGGG GCCCCCCCGT CCCGTCTCCC 15900

10 (2) INFORMATION FOR SEQ ID NO:174:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 414 amino acids  
 (B) TYPE: amino acid  
 15 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

Met Ala Asp Ile Pro Pro Asp Pro Pro Ala Leu Asn Thr Thr Pro Ala  
 1 5 10 15  
 Asn His Ala Pro Pro Ser Pro Pro Pro Gly Ser Arg Lys Arg Arg Arg  
 25 20 25 30  
 Pro Val Leu Pro Ser Ser Ser Glu Ser Glu Gly Lys Pro Asp Thr Glu  
 35 40 45  
 Ser Glu Ser Ser Ser Thr Glu Ser Ser Glu Asp Glu Ala Gly Asp Leu  
 50 55 60  
 30 Arg Gly Gly Arg Arg Arg Ser Pro Arg Glu Leu Gly Gly Arg Tyr Phe  
 65 70 75 80  
 Leu Asp Leu Ser Ala Glu Ser Thr Thr Gly Thr Glu Ser Glu Gly Thr  
 85 90 95  
 Gly Pro Ser Asp Asp Asp Asp Asp Ala Ser Asp Gly Trp Leu Val  
 35 100 105 110  
 Asp Thr Pro Pro Arg Lys Ser Lys Arg Pro Arg Ile Asn Leu Arg Leu  
 115 120 125  
 Thr Ser Ser Pro Asp Arg Arg Ala Gly Val Val Phe Pro Glu Val Trp  
 130 135 140  
 40 Arg Asn Asp Arg Pro Ile Arg Ala Ala Gln Pro Gln Ala Pro Ala Gln  
 145 150 155 160  
 Ser Ser Gly Asp Arg Ala Ala Ala Pro Arg Arg Ser Ala Arg Gln Ala  
 165 170 175

Gln Met Arg Ser Gly Ala Ala Trp Thr Leu Asp Leu His Tyr Ile Arg  
 180 185 190  
 Gln Cys Val Asn Gln Leu Phe Arg Ile Leu Arg Ala Ala Pro Asn Pro  
 195 200 205  
 5 Pro Gly Ser Ala Asn Arg Leu Arg His Leu Val Arg Asp Cys Tyr Leu  
 210 215 220  
 Met Gly Tyr Cys Arg Thr Arg Leu Gly Pro Arg Thr Trp Gly Arg Leu  
 225 230 235 240  
 Leu Gln Ile Ser Gly Gly Thr Trp Asp Val Arg Leu Arg Asn Ala Ile  
 10 245 250 255  
 Arg Glu Val Glu Ala Arg Phe Glu Pro Ala Ala Glu Pro Val Cys Glu  
 260 265 270  
 Leu Pro Cys Leu Asn Ala Arg Arg Tyr Gly Pro Glu Cys Asp Val Gly  
 275 280 285  
 15 Asn Leu Glu Thr Asn Gly Gly Ser Thr Ser Asp Asp Glu Ile Ser Asp  
 290 295 300  
 Ala Thr Asp Ser Asp Asp Thr Leu Ala Ser His Ser Asp Thr Glu Gly  
 305 310 315 320  
 Gly Pro Ser Pro Ala Gly Arg Glu Asn Pro Glu Ser Ala Ser Gly Gly  
 20 325 330 335  
 Ala Ile Ala Ala Arg Leu Glu Cys Glu Phe Gly Thr Phe Asp Trp Thr  
 340 345 350  
 Ser Glu Glu Gly Ser Gln Pro Trp Leu Ser Ala Val Val Ala Asp Thr  
 355 360 365  
 25 Ser Ser Ala Glu Arg Ser Gly Leu Pro Ala Pro Gly Ala Cys Arg Ala  
 370 375 380  
 Thr Glu Ala Pro Glu Arg Glu Asp Gly Cys Arg Lys Met Arg Phe Pro  
 385 390 395 400  
 Ala Ala Cys Pro Tyr Pro Cys Gly His Thr Phe Leu Arg Pro  
 30 405 410

## (2) INFORMATION FOR SEQ ID NO:175:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 287 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 40 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:



	Met	Gly	Val	Val	Val	Val	Ser	Val	Val	Thr	Leu	Leu	Asp	Gln	Arg	Asn
	1				5					10					15	
	Ala	Leu	Pro	Arg	Thr	Ser	Ala	Asp	Asp	Ala	Leu	Trp	Ser	Phe	Leu	Leu
				20					25					30		
5	Arg	Gln	Cys	Arg	Ile	Leu	Ala	Ser	Glu	Pro	Leu	Gly	Thr	Pro	Val	Val
		35					40					45				
	Val	Arg	Pro	Ala	Asn	Leu	Arg	Arg	Leu	Ala	Glu	Pro	Leu	Met	Asp	Leu
		50					55				60					
	Pro	Lys	Phe	Trp	Ile	Val	Arg	Thr	Arg	Ser	Cys	Arg	Cys	Pro	Pro	Asn
10	65					70				75					80	
	Thr	Thr	Thr	Gly	Leu	Phe	Ala	Glu	Asp	Asp	Pro	Leu	Glu	Ser	Ile	Glu
				85					90					95		
	Ile	Leu	Asp	Ala	Pro	Ala	Cys	Phe	Arg	Leu	Leu	His	Gln	Glu	Arg	Pro
				100					105				110			
15	Gly	Pro	His	Arg	Leu	Tyr	His	Leu	Trp	Val	Val	Gly	Ala	Ala	Asp	Leu
			115					120				125				
	Cys	Val	Pro	Phe	Leu	Glu	Tyr	Ala	Gln	Lys	Thr	Arg	Leu	Gly	Phe	Arg
		130						135				140				
	Phe	Ile	Ala	Met	Lys	Thr	Asn	Asp	Ala	Trp	Val	Gly	Glu	Pro	Trp	Pro
20	145					150				155					160	
	Leu	Pro	Asp	Arg	Phe	Leu	Pro	Glu	Arg	Thr	Val	Ser	Trp	Thr	Pro	Phe
				165					170					175		
	Pro	Ala	Ala	Pro	Asn	His	Pro	Leu	Glu	Asn	Leu	Leu	Ser	Arg	Tyr	Glu
				180					185					190		
25	Tyr	Gln	Tyr	Gly	Val	Val	Val	Pro	Gly	Asp	Arg	Glu	Arg	Ser	Cys	Leu
		195						200				205				
	Arg	Trp	Leu	Arg	Ser	Leu	Val	Ala	Pro	His	Asn	Lys	Pro	Arg	Pro	Ala
		210						215				220				
	Ser	Ser	Arg	Pro	His	Pro	Ala	Thr	His	Pro	Thr	Gln	Arg	Pro	Cys	Phe
30	225					230					235				240	
	Thr	Cys	Met	Gly	Arg	Pro	Glu	Ile	Pro	Asp	Glu	Pro	Ser	Trp	Gln	Thr
				245					250					255		
	Gly	Asp	Asp	Asp	Pro	Gln	Asn	Pro	Gly	Pro	Pro	Leu	Ala	Val	Gly	Asp
				260					265				270			
35	Glu	Trp	Pro	Pro	Ser	Ser	His	Val	Cys	Tyr	Pro	Ile	Thr	Asn	Leu	
		275						280					285			

## (2) INFORMATION FOR SEQ ID NO:176:

- 40 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 507 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

Val Gly Gly Cys Val Asp Lys Leu Pro Leu Leu Lys Thr Pro Gly Pro  
 1 5 10 15  
 Val Arg Ala Arg Trp Leu Ala Arg Ala Thr Arg Arg Met Ala Cys Arg  
 10 20 25 30  
 Lys Phe Cys Gly Val Tyr Arg Arg Pro Asp Lys Arg Gln Glu Ala Ser  
 35 40 45  
 Val Pro Pro Glu Thr Asn Thr Ala Pro Ala Phe Pro Ala Ser Thr Phe  
 50 55 60  
 15 Tyr Thr Pro Ala Glu Asp Ala Tyr Leu Ala Pro Gly Pro Pro Glu Thr  
 65 70 75 80  
 Ile His Pro Ser Arg Pro Pro Ser Pro Gly Glu Ala Ala Arg Leu Cys  
 85 90 95  
 Gln Leu Gln Glu Ile Leu Ala Gln Met His Ser Asp Glu Asp Tyr Pro  
 20 100 105 110  
 Ile Val Asp Ala Ala Gly Ala Glu Glu Glu Asp Glu Ala Asp Asp Asp  
 115 120 125  
 Ala Pro Asp Asp Val Ala Tyr Pro Glu Asp Tyr Ala Glu Gly Arg Phe  
 130 135 140  
 25 Leu Ser Met Val Ser Ala Ala Pro Leu Pro Gly Ala Ser Gly His Pro  
 145 150 155 160  
 Pro Val Pro Gly Arg Ala Ala Pro Pro Asp Val Arg Thr Cys Asp Ser  
 165 170 175  
 Gly Lys Val Gly Ala Thr Gly Phe Thr Pro Glu Glu Leu Asp Thr Met  
 30 180 185 190  
 Asp Arg Glu Ala Leu Arg Ala Ile Ser Arg Gly Cys Lys Pro Pro Ser  
 195 200 205  
 Thr Leu Ala Lys Leu Val Thr Gly Leu Gly Phe Ala Ile His Gly Ala  
 210 215 220  
 35 Leu Ile Pro Gly Ser Glu Gly Cys Val Phe Asp Ser Ser His Pro Asn  
 225 230 235 240  
 Tyr Pro His Arg Val Ile Val Lys Ala Gly Trp Tyr Ala Ser Thr Asn  
 245 250 255  
 His Glu Ala Arg Leu Leu Arg Arg Leu Asn His Pro Ala Ile Leu Pro  
 40 260 265 270  
 Leu Leu Asp Leu His Val Val Ser Gly Val Thr Cys Leu Val Leu Pro  
 275 280 285  
 Lys Tyr His Cys Asp Leu Tyr Thr Tyr Leu Ser Lys Arg Pro Ser Pro  
 475

	290		295		300
	Leu Gly His	Leu Gln Ile	Thr Ala Val	Ser Arg Gln	Leu Leu Ser Ala
	305		310		315
	Ile Asp Tyr	Val His Cys	Glu Gly Ile	Ile His Arg	Asp Ile Lys Thr
5		325		330	335
	Glu Asn Ile	Leu Ile Asn	Thr Pro Glu	Asn Ile Cys	Leu Gly Asp Phe
		340		345	350
	Gly Ala Ala	Cys Phe Val	Arg Gly Cys	Arg Ser Ser	Pro Phe His Tyr
		355		360	365
10	Gly Ile Ala	Gly Thr Ile	Asp Thr Asn	Ala Pro Glu	Val Leu Ala Gly
		370		375	380
	Asp Pro Tyr	Thr Gln Val	Ile Asp Ile	Trp Ser Ala	Gly Leu Val Ile
		385		390	395
	Phe Glu Thr	Ala Val His	Thr Ala Ser	Leu Phe Ser	Ala Pro Arg Asp
15		405		410	415
	Pro Glu Arg	Arg Pro Cys	Asp Asn Gln	Ile Ala Arg	Ile Ile Arg Gln
		420		425	430
	Ala Gln Val	His Val Asp	Glu Phe Pro	Thr His Ala	Glu Ser Arg Leu
		435		440	445
20	Thr Ala His	Tyr Arg Ser	Arg Ala Ala	Gly Asn Asn	Arg Pro Ala Trp
		450		455	460
	Trp Ala Trp	Thr Arg Tyr	Tyr Lys Ile	His Thr Asp	Val Glu Tyr Leu
		465		470	475
	Ile Cys Lys	Ala Leu Thr	Phe Asp Ala	Ala Leu Arg	Pro Ser Ala Ala
25		485		490	495
	Glu Leu Leu	Arg Leu Pro	Leu Phe His	Pro Lys	
		500		505	

## (2) INFORMATION FOR SEQ ID NO:177:

30

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 392 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

35

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

40

Val Cys Ile	Ala Tyr His	Gly Met Gly	Arg Leu Thr	Ser Gly Val	Gly
1		5		10	15
Thr Ala Ala	Leu Leu Val	Val Ala Val	Gly Leu Arg	Val Val Cys	Ala
					476

477

Ala Pro Lys Arg Leu Arg Leu Pro His Ile Arg Asp Asp Asp Ala Pro  
 370 375 380  
 Pro Ser His Gln Pro Leu Phe-Tyr  
 385 390

5

## (2) INFORMATION FOR SEQ ID NO:178:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 392 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

## (ii) MOLECULE TYPE: peptide

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

Val Cys Ile Ala Tyr His Gly Met Gly Arg Leu Thr Ser Gly Val Gly  
 1 5 10 15  
 20 Thr Ala Ala Leu Leu Val Val Ala Val Gly Leu Arg Val Val Cys Ala  
 20 25 30  
 Lys Tyr Ala Asp Pro Ser Leu Lys Met Ala Asp Pro Asn Arg Phe Arg  
 35 40 45  
 Gly Lys Asn Leu Pro Val Leu Asp Gln Leu Thr Asp Pro Pro Gly Val  
 25 50 55 60  
 Lys Arg Val Tyr His Ile Gln Pro Ser Leu Glu Asp Pro Phe Gln Pro  
 65 70 75 80  
 Pro Ser Ile Pro Ile Thr Val Tyr Tyr Ala Val Leu Glu Arg Ala Cys  
 85 90 95  
 30 Arg Ser Val Leu Leu His Ala Pro Ser Glu Ala Pro Gln Ile Val Arg  
 100 105 110  
 Gly Ala Ser Asp Glu Ala Arg Lys His Thr Tyr Asn Leu Thr Ile Ala  
 115 120 125  
 Trp Tyr Arg Met Gly Asp Asn Cys Ala Ile Pro Ile Thr Val Met Glu  
 35 130 135 140  
 Tyr Thr Glu Cys Pro Tyr Asn Lys Ser Leu Gly Val Cys Pro Ile Arg  
 145 150 155 160  
 Thr Gln Pro Arg Trp Ser Tyr Tyr Asp Ser Phe Ser Ala Val Ser Glu  
 165 170 175  
 40 Asp Asn Leu Gly Phe Leu Met His Ala Pro Ala Phe Glu Thr Ala Gly  
 180 185 190  
 Thr Tyr Leu Arg Leu Val Lys Ile Asn Asp Trp Thr Glu Ile Thr Gln  
 195 200 205

Phe Ile His Arg Ala Arg Ala Ser Cys Lys Tyr Ala Leu Pro Leu Arg  
 210 215 220  
 Ile Pro Pro Ala Ala Cys Leu Thr Ser Lys Ala Tyr Gln Gln Gly Val  
 225 230 235 240  
 5 Thr Val Asp Ser Ile Gly Met Leu Pro Arg Phe Ile Pro Glu Asn Gln  
 245 250 255  
 Arg Thr Val Ala Lys Leu Lys Ile Ala Gly Trp His Gly Pro Lys Pro  
 260 265 270  
 Pro Tyr Thr Ser Thr Leu Leu Pro Pro Glu Leu Ser Asp Thr Thr Asn  
 10 275 280 285  
 Ala Thr Gln Pro Glu Leu Val Pro Glu Asp Pro Glu Asp Ser Ala Leu  
 290 295 300  
 Leu Glu Asp Pro Ala Gly Thr Val Ser Ser Gln Ile Pro Pro Asn Trp  
 305 310 315 320  
 15 His Ile Pro Ser Ile Gln Asp Val Ala Pro His His Ala Pro Ala Ala  
 325 330 335  
 Pro Ser Asn Pro Gly Leu Ile Ile Gly Ala Gly Ser Thr Leu Ala Val  
 340 345 350  
 Leu Val Ile Gly Gly Ile Ala Phe Trp Val Arg Arg Arg Ala Gln Met  
 20 355 360 365  
 Ala Pro Lys Arg Leu Arg Leu Pro His Ile Arg Asp Asp Asp Ala Pro  
 370 375 380  
 Pro Ser His Gln Pro Leu Phe Tyr  
 385 390

25

## (2) INFORMATION FOR SEQ ID NO:179:

## (i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 429 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

Val Tyr Leu Trp Ala Arg Val Gly Gly Trp Leu Gly Tyr Leu Gly Gly  
 1 5 10 15  
 40 Thr Trp Thr Pro His Lys Gly Ser Leu Glu Gly Gly Lys Leu Gly Gln  
 20 25 30  
 Phe Ile Gly Arg Glu Arg Gly Ala Arg Thr Ala Val Pro Thr Ile Ser  
 35 40 45

His Arg Ala His Ser His Leu Asp Pro Ser Asp Pro Gly Met Pro Gly  
 50 55 60  
 Arg Ser Leu Gln Gly Leu Ala Ile Leu Gly Leu Trp Val Cys Ala Thr  
 65 70 75 80  
 5 Gly Leu Val Val Arg Gly Pro Thr Val Ser Leu Val Ser Asp Ser Leu  
 85 90 95  
 Val Asp Ala Gly Ala Val Gly Pro Gln Gly Phe Val Glu Glu Asp Leu  
 100 105 110  
 Arg Val Phe Gly Glu Leu His Phe Val Gly Ala Gln Val Pro His Thr  
 10 115 120 125  
 Asn Tyr Tyr Asp Gly Ile Ile Glu Leu Phe His Tyr Pro Leu Gly Asn  
 130 135 140  
 His Cys Pro Arg Val Val His Val Val Thr Leu Thr Ala Cys Pro Arg  
 145 150 155 160  
 15 Arg Pro Ala Val Ala Phe Thr Leu Cys Arg Ser Thr His His Ala His  
 165 170 175  
 Ser Pro Ala Tyr Pro Thr Leu Glu Leu Gly Leu Ala Arg Gln Pro Leu  
 180 185 190  
 Leu Arg Val Arg Thr Ala Thr Arg Asp Tyr Ala Gly Val Leu Arg Val  
 20 195 200 205  
 Trp Val Gly Ser Ala Thr Asn Ala Ser Leu Phe Val Leu Gly Val Ser  
 210 215 220  
 Ala Asn Gly Thr Phe Val Tyr Asn Gly Ser Asp Tyr Gly Ser Cys Asp  
 225 230 235 240  
 25 Pro Ala Gln Leu Pro Phe Ser Ala Pro Arg Leu Gly Pro Ser Ser Val  
 245 250 255  
 Tyr Thr Pro Gly Ala Ser Arg Pro Thr Pro Pro Arg Thr Thr Thr Ser  
 260 265 270  
 Pro Ser Ser Pro Arg Asp Pro Thr Pro Ala Pro Gly Asp Thr Gly Thr  
 30 275 280 285  
 Pro Ala Pro Ala Ser Gly Glu Arg Ala Pro Pro Asn Ser Thr Arg Ser  
 290 295 300  
 Ala Ser Glu Ser Arg His Arg Leu Thr Val Ala Gln Val Ile Gln Ile  
 305 310 315 320  
 35 Ala Ile Pro Ala Ser Ile Ile Ala Phe Val Phe Leu Gly Ser Cys Ile  
 325 330 335  
 Cys Phe Ile His Arg Cys Gln Arg Arg Tyr Arg Arg Pro Arg Gly Gln  
 340 345 350  
 Ile Tyr Asn Pro Gly Gly Val Ser Cys Ala Val Asn Glu Ala Ala Met  
 40 355 360 365  
 Ala Arg Leu Gly Ala Glu Leu Arg Ser His Pro Asn Thr Pro Pro Lys  
 370 375 380  
 Pro Arg Arg Arg Ser Ser Ser Ser Thr Thr Met Pro Ser Leu Thr Ser  
 480

385                      390                      395                      400  
 Ile Ala Glu Glu Ser Glu Pro Gly Pro Val Val Leu Leu Ser Val Ser  
                          405                      410                      415  
 Pro Arg Pro Arg Ser Gly Pro Thr Ala Pro Gln Glu Val  
 5                      420                      425

## (2) INFORMATION FOR SEQ ID NO:180:

## (i) SEQUENCE CHARACTERISTICS:

- 10            (A) LENGTH: 430 amino acids  
               (B) TYPE: amino acid  
               (C) STRANDEDNESS: single  
               (D) TOPOLOGY: linear

- 15            (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

Met Arg Ala Gly Leu Val Phe Phe Val Gly Val Trp Val Val Ser Cys  
 20        1                      5                      10                      15  
 Leu Ala Ala Ala Pro Arg Thr Ser Trp Lys Arg Val Thr Ser Gly Glu  
                          20                      25                      30  
 Asp Val Val Leu Leu Pro Ala Pro Ala Gly Pro Glu Glu Arg Thr Arg  
                          35                      40                      45  
 25        Ala His Lys Leu Leu Trp Ala Ala Glu Pro Leu Asp Ala Cys Gly Pro  
                          50                      55                      60  
 Leu Arg Pro Ser Trp Val Trp Pro Pro Arg Arg Val Leu Glu Thr Val  
 65                                   70                                   75                                   80  
 Val Asp Ala Ala Cys Met Arg Ala Pro Glu Pro Leu Ala Ile Ala Tyr  
 30                                   85                                   90                                   95  
 Ser Pro Pro Phe Pro Ala Gly Asp Glu Gly Ser Glu Leu Ala Trp Arg  
                          100                      105                      110  
 Asp Arg Val Ala Val Val Asn Glu Ser Leu Val Ile Tyr Gly Ala Leu  
                          115                      120                      125  
 35        Glu Thr Asp Ser Gly Thr Leu Ser Val Val Gly Leu Ser Asp Glu Ala  
                          130                      135                      140  
 Arg Gln Val Ala Ser Val Val Leu Val Val Glu Pro Ala Pro Val Pro  
 145                                   150                                   155                                   160  
 Thr Pro Thr Pro Asp Asp Tyr Asp Glu Glu Asp Asp Ala Gly Val Ser  
 40                                   165                                   170                                   175  
 Thr Pro Val Ser Val Pro Pro Pro Thr Pro Pro Arg Arg Pro Pro Val  
                          180                      185                      190  
 Ala Pro Pro Thr His Pro Arg Val Ile Pro Glu Val Ser His Val Arg



	195	200	205
	Gly Val Thr Val His Met Pro Glu Ala Ile Leu Phe Ala Pro Gly Glu		
	210	215	220
	Thr Phe Gly Thr Asn Val Ser Ile His Ala Ile Ala His Asp Asp Gly		
5	225	230	235 240
	Pro Tyr Ala Met Asp Val Val Trp Met Arg Phe Asp Val Pro Ser Ser		
	245	250	255
	Cys Ala Glu Met Arg Ile Tyr Glu Ala Cys Leu Tyr His Pro Gln Leu		
	260	265	270
10	Pro Glu Cys Leu Ser Pro Ala Asp Ala Pro Cys Ala Val Ser Ser Trp		
	275	280	285
	Ala Tyr Arg Leu Ala Val Arg Ser Tyr Ala Gly Cys Ser Arg Thr Thr		
	290	295	300
	Pro Pro Pro Arg Cys Phe Ala Glu Ala Arg Met Glu Pro Val Pro Gly		
15	305	310	315 320
	Leu Ala Trp Leu Ala Ser Thr Val Asn Leu Glu Phe Gln His Asp Gln		
	325	330	335
	His Ala Gly Leu Cys Val Val Tyr Val Asp Asp His Ile His Ala Trp		
	340	345	350
20	Gly His Met Thr Ile Ser Thr Ala Ala Gln Tyr Arg Asn Ala Val Val		
	355	360	365
	Glu Gln His Leu Pro Gln Arg Gln Pro Glu Pro Val Glu Pro Trp His		
	370	375	380
	Val Arg Ala Pro Pro Pro Ala Pro Ser Arg Pro Leu Arg Leu Gly Ala		
25	385	390	395 400
	Val Leu Gly Ala Ala Leu Leu Leu Ala Ala Leu Gly Leu Ser Ala Trp		
	405	410	415
	Gly Val His Asp Leu Leu Ala Gln Ala Leu Leu Ala Gly Gly		
	420	425	430

30

(2) INFORMATION FOR SEQ ID NO:181:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 41 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

Val His Ala Val Asp Ala Pro Ser Gln Phe Val Thr Trp Leu Ala Val

	1		5		10		15
	Arg	Trp	Leu	Arg	Gly	Ala	Val
					Gly	Leu	Gly
					Ala	Val	Leu
						Cys	Gly
							Ile
			20		25		30
	Ala	Phe	Tyr	Val	Thr	Ser	Ile
5						Arg	Ala
		35		40			

## (2) INFORMATION FOR SEQ ID NO:182:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 85 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 15 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

	Met	Thr	Ser	Arg	Pro	Ala	Asp	Gln	Asp	Ser	Val	Arg	Ser	Ser	Ala	Ser
20	1				5				10					15		
	Val	Pro	Leu	Tyr	Pro	Ala	Asp	Val	Pro	Ala	Glu	Ala	Tyr	Tyr	Ser	Glu
				20				25						30		
	Ser	Glu	Asp	Glu	Ala	Ala	Asn	Asp	Phe	Leu	Val	Arg	Met	Gly	Arg	Gln
			35				40					45				
25	Gln	Ser	Val	Leu	Arg	Arg	Arg	Arg	Arg	Thr	Arg	Cys	Val	Gly	Leu	
		50				55				60						
	Val	Ile	Ala	Cys	Leu	Val	Val	Leu	Ser	Gly	Gly	Phe	Gly	Ala	Leu	Leu
	65				70					75					80	
	Val	Trp	Leu	Leu	Arg											
30					85											

## (2) INFORMATION FOR SEQ ID NO:183:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 296 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 40 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

Met Ile Arg Arg Arg Gly Asn Val Glu Ile Arg Val Tyr Tyr Glu Ser  
 1 5 10 15  
 Val Arg Pro Ser Arg Ser Arg Ser His Leu Lys Pro Ser Asp His Gln  
 20 25 30  
 5 Glu Phe Pro Gly His His Val Ser Pro Gly Ser Pro Gly Phe Pro Glu  
 35 40 45  
 Ser Pro Gly Asn Arg Glu Phe His Asp Leu Pro Glu Asn Pro Gly Ser  
 50 55 60  
 Arg Ala Tyr Pro Gly Thr Arg Asp Pro His Asp Pro His Gly Cys Pro  
 10 65 70 75 80  
 Gly Ser Leu Asp Pro His Gly Asn Pro Ala Gln Pro Ala Gly Leu Pro  
 85 90 95  
 Ser Pro Val Pro Tyr Ala Pro Leu Gly Ser Pro Asp Pro Ser Ser Pro  
 100 105 110  
 15 Arg Gln Arg Thr Tyr Val Leu Pro Arg Val Gly Ile Arg Asn Ala Pro  
 115 120 125  
 Ala Ser Asp Thr Arg Ala Pro Lys Arg Ala His Ser Arg His Arg Ala  
 130 135 140  
 Asp Arg Pro Pro Glu Ser Pro Gly Ser Glu Leu Tyr Pro Leu Asn Ala  
 20 145 150 155 160  
 Gln Ala His Leu Gln Met Leu Pro Ala Asp His Arg Ala Phe Phe Arg  
 165 170 175  
 Thr Val Ile Glu Val Ser Arg Leu Cys Ala Leu Asn Thr His Asp Pro  
 180 185 190  
 25 Pro Pro Pro Leu Ala Gly Ala Arg Val Gly Gln Glu Ala Gln Leu Val  
 195 200 205  
 His Thr Gln Trp Leu Arg Ala Asn Arg Glu Ser Ser Pro Leu Trp Pro  
 210 215 220  
 Trp Arg Thr Ala Ala Met Asn Phe Ile Ala Ala Ala Ala Pro Cys Val  
 30 225 230 235 240  
 Gln Thr His Met His Asp Leu Leu Met Ala Cys Ala Phe Trp Cys Cys  
 245 250 255  
 Leu Ala His Ala Ser Thr Cys Ser Tyr Ala Gly Ser Ala His Cys Gln  
 260 265 270  
 35 His Leu Phe Arg Ala Phe Gly Cys Gly Pro Pro Val Leu Thr Thr Ser  
 275 280 285  
 Arg Gly Gln Gly Gly Trp Cys Asn  
 290 295

40

(2) INFORMATION FOR SEQ ID NO:184:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 178 base pairs

484

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

	ATGCACGTGT AACC GCCAGT CCGTGCTTGC CTAGCGAACT CACCCGTCCC GGCTGGCGTG	60
	CGCAGCCCCG GCCGTGTTGC GGGCCCTCTT AAGGGGCGGC GGCAGGACGG GGA CTCCGCC	120
10	CCGCCTCCTT TCCCCGGGG AGTCAACCCC CGGGGGGGTG TATTCTGGGG GGGGGGT	178

(2) INFORMATION FOR SEQ ID NO:185:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 2116 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

	GACCCAGATC CCACCCCGC CCGCAACGGG GCGCCGCCGC TGCTGCTGCT CCGCGGGGCG	60
	CCAGGGGGCG CCGGTCGGGT CGCGGCGGGC TGGGAGGTTC CGCGGGTCGC CCCC GCACCG	120
25	CCGCCCCCGC GCCGGGGCGC TCTTCGGGGG GCGGGCGGGA CGTAGTCCGC TGCAGAGGGA	180
	GACAGAGACG GGAACCCCGG GTTAGTGCCN GACCCCGGCC CGACCCCGC CCAGTGCCCG	240
	ACCCCGGCC GACCCCGGCC CGACCCCGC CCAGTGCCCG ACCCCCGCC AGTGCCCGAC	300
	CCCCGCCAG TGCCCGACCC CCGCCAGTG CCGACCCCG CCGCGCCCTC ACCGTGCGGC	360
	AGGTCACTCGT CCTCGTCGTC CGTGCCGGGC CACGGGGGGG TGGGCGACAG GCGCGGACC	420
30	GTGTGTCCCC CCAGCGACAG GGAGCGCGGG GCCGTCCGCG GGTTGCCCCG CCAGATAAAG	480
	TCCACGGCCG TGCCGGACCG CACGGCCGCC TCGGCCTCCA CGCGGGTCCG GGGGTGCTTC	540
	ACTATCGGGA TGGTGCTGAA CGACCCGCTG GCGGTCACGC CCACTATCAG GTACGCCACC	600
	GGGTGTGTTG ACAGGGGACA CGTGTGCGC AACGGAATCC AGGTCTTCAT GCACGGGATG	660
	CAGAAGGGGT GCAGGCAGGG AAAACTCTGG CAGCGCAGGG GCGGGGCGAT CTCGTCCGTG	720
35	CACACGGCAC ACACGTCGCC CCCCCCTCCC GCTTCCGCTT CCTCCTCACC CACGGGCCCCA	780
	CCCCCGCAGG ATCCCTGCGC GTCGGCGGGC GTGGGGCTGC CCTGGCGCTC GGCCGGGGGC	840
	CGGGCCGGGG GCGTGGCCGC GTCCATCAGG CCCGCCTCGA ACATCTCCGT GTCCGTGCTG	900
	CCCGCCTCGG AGGTGGAGTC GCGGTGAAGG TCGTCGTCAG AGATTCCCAC CTCGGTCTCC	960
	TCCTCCGAGT CGCTGCTGGC GAGCCACTGC ATGTCGTTGA GCATCCCCCA GCGGTGCGGG	1020
40	GCGGCGGGCT GCTTGACAAA GAAACGGGGG GGGATTAGA GGGCGCGGGG CGTGAGGCGG	1080
	GACCCCGGTG CCGTGTCCCC CGTGTCCCTC CCTCACCCCG GCGCCCGGCC CGCTGCTTTT	1140
	TGTTCCGAAG GGGGGGAGAA AGGGGTCCGT AACCAAAGGT GGTCTGCGTC CTTTGATTC	1200
	CGACCCCTCG TCTCCCCCCC CCTGTCCCC GCTCTCGGGC TCAGGGCTCC CTGCCTCCCT	1260

```

CGCCCCCCCA AAGGGTCGGG GGGCGGCGCA CGGCCCCACGG GGGTCCCCCG ACCGCTTAAG 1320
CGGGCCGGGG GTCGGCCCCG TCAAGCGTCC CCGCCCCCGA GCCCACC GCCGACCACC 1380
CCCAACCCGC AGCCGGGTGG TCCGGGGAAA AGGGGGGGCC TGAGACCCGG GGGTCGCCCT 1440
CTCACCGTGC CGGGGTCTG CCGCGGCGGC CGCTCGGGGC CGGGGTCCGC CCGGGAGCTC 1500
5 GTGCCGGGCC GGGGTTCAT GAGCCGGGT AGGGTAGACT CGAGACGGC GCCCGCGGTC 1560
TCTCTTTGC CGGGTGTAG TCTCTGTCTC TCCGGGTCTC CTCCTCCCGC CGGGCCGCCG 1620
CTCCGTCGCT CGCAGTGCCG GGGTGCGAAT GCGGCCCCGAC CGTCACACGG GGCTGCCTTA 1680
TACCCGGCGC CTATCCACTC CCCCAGAGG GCGGCATTTA CGATTCCCCC AATAGCCGCG 1740
CGCCCCGGCG GGGGCGGAGG GAGGGAATCC CCCCCTCTCG GGGCGGCCCC GTCCCCGGG 1800
10 ACCAACCGGG TGTACTCAA GAACCCATT AGCATGCGCC GCCCCCGCC GACGCAGATG 1860
GGAGTCCCCC CGGCGCCCCG CCGGCGCGGC CCTGAGTGGT GCGCGCCCCC GGGGAGAAAT 1920
TCATTAGCAT ACTAGGAAGC CCAGGGGACC AATAGGGGCC GATCAGCCCA CCCACCCGGC 1980
GGCGCGCGAG GCTCTGCGTG TTCTGCCAAG AAAGTAATCA GCATAACCCG GAACCCCGAG 2040
GGAGTAATTA CGCGGGGAGC GAGGGGCCGT CCGAACGTTT TTAATTACCA TAAGCGGGAA 2100
15 TGGCGGCCCG TTAAA 2116

```

## (2) INFORMATION FOR SEQ ID NO:186:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 338 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 25 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

```

Met Leu Asn Asp Met Gln Trp Leu Ala Ser Ser Asp Ser Glu Glu Glu
30 1 5 10 15
Thr Glu Val Gly Ile Ser Asp Asp Asp Leu His Arg Asp Ser Thr Ser
20 25 30
Glu Ala Gly Ser Thr Asp Thr Glu Met Phe Glu Ala Gly Leu Met Asp
35 35 40 45
Ala Ala Thr Pro Pro Ala Arg Pro Pro Ala Glu Arg Gln Gly Ser Pro
50 55 60
Thr Pro Ala Asp Ala Gln Gly Ser Cys Gly Gly Pro Val Gly Glu
65 70 75 80
Glu Glu Ala Glu Ala Gly Gly Gly Gly Asp Val Cys Ala Val Cys Thr
40 85 90 95
Asp Glu Ile Ala Pro Pro Leu Arg Cys Gln Ser Phe Pro Cys Leu His
100 105 110
Pro Phe Cys Ile Pro Cys Met Lys Thr Trp Ile Pro Leu Arg Asn Thr

```

115 120 125  
 Cys Pro Leu Cys Asn Thr Pro Val Ala Tyr Leu Ile Val Gly Val Thr  
 130 135 140  
 Ala Ser Gly Ser Phe Ser Thr Ile Pro Ile Val Asn Asp Pro Arg Thr  
 5 145 150 155 160  
 Arg Val Glu Ala Glu Ala Ala Val Arg Ser Gly Thr Ala Val Asp Phe  
 165 170 175  
 Ile Trp Thr Gly Asn Pro Arg Thr Ala Pro Arg Ser Leu Ser Leu Gly  
 180 185 190  
 10 Gly His Thr Val Arg Ala Leu Ser Pro Thr Pro Pro Trp Pro Gly Thr  
 195 200 205  
 Asp Asp Glu Asp Asp Asp Leu Ala Asp Gly Glu Gly Gly Arg Gly Ser  
 210 215 220  
 Gly Thr Gly Arg Gly Ser Gly Thr Gly Arg Gly Ser Gly Thr Gly Arg  
 15 225 230 235 240  
 Gly Ser Gly Thr Gly Arg Gly Ser Gly Gly Gly Arg Ala Gly Val Gly  
 245 250 255  
 His Trp Ala Gly Val Gly Arg Gly Xaa Gly Thr Asn Arg Gly Phe Pro  
 260 265 270  
 20 Ser Leu Ser Pro Ser Ala Ala Asp Tyr Val Pro Pro Ala Pro Arg Arg  
 275 280 285  
 Ala Pro Arg Arg Gly Gly Gly Gly Ala Gly Ala Thr Arg Gly Thr Ser  
 290 295 300  
 Gln Pro Ala Ala Trp Ala Pro Pro Gly Ala Pro Arg Ser Ser Ser Ser  
 25 305 310 315 320  
 Gly Gly Ala Pro Leu Arg Ala Gly Val Gly Ser Gly Ser Xaa Xaa Xaa  
 325 330 335  
 Xaa Xaa

30

(2) INFORMATION FOR SEQ ID NO:187:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 642 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

CGCGGCGGTT TCCGCGTTCC GTTCTCTCTC CCTCCCGGCC GCCCGCTCC CGGGCCCGAC 60  
 CCTCGCCCCT TCCCTTCTCC TCGTCTTCCC CCGTCCCGGCC GCGCCCTTC CCTCTTCTT 120

CTCTCTCTCT GTCTCGCTGT CTCGCTCTCC TCACATTTCC CCCCCCCCCC CCCGCCGCCG 180  
 CCGCCGCCCT CTGCCCCGCT CCCACCGAGA CGCCGCGCCG CGTGAGCCGT CCGCCGGGGG 240  
 ACCCAGGCTC CGGGGGGGGG GCGCGCCTGC GTGTGTCTCG TGTGAGAGAG CGCGCCCCCTC 300  
 GAACGCCGCG CGTTCTCGCA GGTAGGTTTA GGGTCGTACA GGTGAGCTTC TGCTGAGGCG 360  
 5 GCGGGAGAGG GGGGGCGGG CGGAAGAGAG AAGAGAGCAG GGGTTGGGGG AAAACTGTTC 420  
 TTCCTCCCC TTTCAAGAAA CACGAGGCGG GGGTCCCAGA AAGGGCAGGC AGGTCAGCCG 480  
 CACCGCCCGC GAGCCAACCC GTATCCTTTT TTTCTAGGTG TTTTGTGTTT TGTTCCTGTT 540  
 TTTGTTTGT TGTATTAT TTTCGCGGAT CCGGCGTGT CGGATCCACC CCCCCCTTC 600  
 TCCTTCCTCT TCCCTTCCAC CCACCCCGT TCCCCCCCC C 642

10

(2) INFORMATION FOR SEQ ID NO:188:

## (i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 353 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

CGCGCCCCCG CCCGGCCGCC GCGCGCCCCC GCGCGACCGC CGCGCGCCCC CGCCCGGCCG 60  
 CCGCGCGCCC CCGCCCGGCC GCGCGCTCG CGCCGCGCC CCTCCCGGC GCTTCCGGGG 120  
 CCTTTCCTTC CTCCCCGCC GCGACCCCGG CCGCGCCCA CCGCCCGCC CGGCAGGGGG 180  
 25 GCGCGCGCGC CGCGCAGAAC ACACAGACGA ACACACGGTG GCGATCTTTT CTTTACTTCG 240  
 GCAGACCAGC GAGCCCCGGC CCCGGCCCGC GCGCGCCGC CACACCCACG GCACCCCCC 300  
 CGCGCCAC CCCGGGTCC ACACAGGAGC GCGCGGCGG CAGAAACGCG GG 353

30

(2) INFORMATION FOR SEQ ID NO:189:

## (i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 6386 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

40 CACGCGCCGC ATTTGCGCCC GGGGCCCCGC GCTCCCCCG GCGGCCTGG CCGTCGGGG 60  
 CCAGATGTAC GTGAACCGCA ACGAGATCTT CAACGCCGCG CTGGCCGTTA CGAACATCAT 120  
 CCTGGATCTG GACATCGCCC TGAAGGAGCC CGTCCCCCTT CCGGCTCC ACGAGGCCCT 180  
 GGGTCACTTT AGGCGCGGGG CGCTGGCGGC GGTTCAGCTG TTGTTTCCCG CGGCCCGCT 240

	AGACCCCGAC	GCCTATCCCT	GTTATTTTTT	CAAAAGCGCC	TGTCGGCCCC	GCGCGCCGCC	300
	CGTCTGTGCG	GGCGACGGGC	CCTCGGCCGG	TGGCGACGAC	GGCGACGGGG	ACTGGTTCCC	360
	CGACGCCCGGT	GGCGACGACG	GCGACGAGGA	GTGGGAGGAG	GACACGGACC	CCATGGACAC	420
	GACCCACGGC	CCCCTCCCCG	ACGACGAGGC	CGCGTACCTC	GACCTGCTAC	ACGAACAGAT	480
5	ACCAGCGGCG	ACGCCCAGCG	AACCGGACTC	CGTCGTGTGT	TCCTGCGCCG	ACAAGATCGG	540
	GCTGCGCGTG	TGCCTACCGG	TCCCCGCCCC	GTACGTTGTG	CACGGCTCCC	TGACGATGCG	600
	TGGGGTGGCG	AGGGTGATCC	AGCAGGCGGT	GCTGTTGGAC	CGCGACTTCG	TGGAGGCCGT	660
	AGGGAGCCAC	GTAAAGAACT	TTTTGCTGAT	CGATACGGGC	GTGTACGCCC	ACGGCCACAG	720
	CCTGCGCTTG	CCGTATTTTC	CCAAGATCGG	CCCCGACGGC	TCCGCGTGCG	GCCGGTTATT	780
10	GCCCGTCTTC	GTGATCCCCC	CCGCGTGCGA	GGACGTTCCG	GCGTTCGTCTG	CCGCGCACGC	840
	CGACCCCGGG	CGCTTCCACT	TTCACGCCCC	GCCCATGTTT	TCCGCGGCCC	CGCGGGAGAT	900
	CCGCGTCTTC	CACAGCCTGG	GCGGGGACTA	TGTCAGCTTT	TTCGAGAAGA	AGGCGTCGCG	960
	CAACGCCCTG	GAGCACTTTG	GCGGACGCGA	GACCCTGACG	GAGGTTCTGG	GCCGCTACGA	1020
	TGTGCGGCCC	GACGCCGGGG	AGACCGTGGA	GGGGTTCGCG	TCAGAACTGC	TGGGGCGAAT	1080
15	AGTCGCGTGC	ATCGAGGCCC	ACTTTCGCCA	GCACGCGCGG	GAATATCAGG	CCGTGTCCGT	1140
	TCGCCGGGCC	GTCAATTAAG	ACGACTGGGT	CCTGCTGCAG	CTGATCCCCG	GCCGCGGCGC	1200
	CCTGAACCAA	AGCCTCTCGT	GTCTGCGCTT	CAAGCACGGC	AGGGCAAGTC	GCGCGACGGC	1260
	CCGACCTTTT	CTCGCGCTGA	GCGTCGGGAC	CAACAACCGC	CTATGCGCGT	CCCTGTGTCA	1320
	GCAGTGCTTT	GCCACTAAAT	GCGATAACAA	CCGCCTGCAC	ACGCTGTTTA	CCGTGATGTC	1380
20	GGGCACGCCA	TGCTCGCGGT	CCGCTCCCTC	CAGCACCTCA	CGACCGTCAT	CTTCATAACG	1440
	GCCTACGGCC	TCGTGCTCGC	GTGGTACATC	GTCTTTGGTG	CCAGTCCGCT	CCACCGATGT	1500
	ATTTACGCGG	TGCGCCCCGC	CGGGGCACAC	AACGATAACG	CCCTCGTGTG	GATGAAGATA	1560
	AACCAGACGC	TGTTGTTTCT	GGGCCCCCGG	ACCGCCCCCC	CCGGCGGGGC	ATGGACCCCC	1620
	CACGCCACAG	TCGTCTACGC	CAATATCATC	GAAGGTCGGG	CCGTGTCCCT	CCCGGCCATC	1680
25	CCCGCGGCCA	TGAGCCGCCG	GGTCATGAAC	GTGCACGAGG	CCGTAAACTG	CTTGAGAGCC	1740
	CTCTGGGACA	CCCAGATGCG	CCTGGTGGTC	GTCGGTTGGT	TTCTGTATCT	AGCGTTTCGT	1800
	GCCCTTCACC	AACGACGATG	CATGTTGCGC	GTCGTGAGTC	CCGCGCACAG	CATGGTGGCC	1860
	CCGCGACCTT	ATCTTTTGAA	CTACGCCGGC	CGCATAGTGT	CGAGCGTGTT	CTTGCAATAC	1920
	CCCTACACGA	AAATCACCCG	CCTCCTCTGC	GAGCTATCCG	TTCAACGCCA	GACCCTGGTG	1980
30	CAGCTGTTTC	AGGCGGATCC	GGTCACCTTC	TTGTACCACC	GCCCGGCCGT	TGGCGTCATC	2040
	GTGGGCTGCG	AGCTGCTGCT	CCGCTTCGTG	GCCCTCGGTC	TCATCGTCCG	CACCGCTCTC	2100
	ATCTCCCGGG	GCGCCTGCGC	GATCACATAC	CCCCGTGTTT	TAACAATCAC	CACCTGGTGT	2160
	TTCTGTGCCA	TCATCGCCCT	GACGGAGCTG	TATTTTCATC	TGCGGCGGGA	CTCGGCCCCC	2220
	AAAAACGCGG	AACCAGCGGC	CCCCAGGGGG	CGCTCCAAAG	GGTGGTCGGG	CGTCTGCGGG	2280
35	CGCTGCTGTT	CCATCATCCT	CTCCGGTATC	GCCGTGCGCC	TGTGCTATAT	CGCCGTCGTG	2340
	GCCGGGGTGG	TGCTTATGGC	GCTTCGCTAC	GAACAGGAGA	TTCAGCGGCG	CCTGTTTGAT	2400
	CTGTGACGTA	ACGCCTCTTC	CGTTGGAAGA	GGCGGACCCA	GTCGCCCATG	CAAATTAAAT	2460
	ACACGACCCG	CCTCGGGCCT	ACGCACCTTC	GCACGTGCGA	TGCAAATTAA	AATCGTGCAC	2520
	AGAGCCGATC	CGGCCTCGGG	TCTGCTTGCC	CCTCCCCCGG	TCCAGCACAG	GCAGGCTCGT	2580
40	CCGACTTCCG	CATACACCCC	ACCCTACCGC	GTGCTTCCGC	ACCCCCGCCT	ACGCGTGATC	2640
	GCGAAGGCGG	ACCCAGACCT	GCCGTATGCT	AATTAAATAC	ATAAAACCCA	CCCTCGGCGT	2700
	CCGATTGGTT	TCTGGGGACG	GCGGGGGCGG	GGGCGGTGAC	GCCCGACGGG	GAGGGACAAG	2760
	GAGGAGTTTC	GGAAAGCCGG	CCCCGGTCGT	GCGGGTATAA	GGGCAGCCAC	CGGCCCCACTG	2820



	GGCGCTGTGT GCTGCCGTGT GCCGACCCCG GTTGCGCGTC GGTGCCGCTC CTCGATTCCG	2880
	ACCCGGCCAC TCTCTTCCGA CACGCGCCCC CTCGGAGGAC ACCCGCCATC CCAGCCCCGG	2940
	CGACCTACAA CATGGCTACC GACATTGATA TGCTAATCGA CCTAGGATTG GACCTGTCCG	3000
	ACAGCGAGCT CGAGGAGGAC GCTCTGGAGC GGGACGAGGA GGGCCGCCGC GACGACCCCG	3060
5	AGTCCGACAG CAGCGGGGAG TGTTCCTCGT CGGACGAGGA CATGGAAGAC CCCTGCGGAG	3120
	ACGGAGGGGC GGAGGCCATC GACGCGGCGA TTCCCAAAGG TCCCCCGGCC CGCCCCGAGG	3180
	ACGCCGGCAC CCCCGAAGCC TCGACGCCCTC GCCCGGCAGC GCGCGGGGA GCCGACGATC	3240
	CGCCACCCGC GACCACCGGC GTGTGGTCGC GCCTCGGGAC CAGGCGGTCG GCTTCCCCC	3300
	GGGAACCGCA CGGGGGGAAG GTGGCCCGCA TCCAACCCCC GTCGACCAAG GCACCGCATC	3360
10	CCCAGGCGG GCGGCGAGGT CGCCGCCGGG GCCGGGGTCG ATACGGCCCC GCGGCGCGCG	3420
	ACTCCACACC AAACCCCGC CGGCGCGTCT CCAGAAACGC CCACAACCAA GGGGGTCGCC	3480
	ACCCCGCGTC GCGCGGACG GACGGCCCCG GCGCCACCCA CGGCGAGGCG CGGCGCGGAG	3540
	GGGAGCAGCT CGACGTCTCC GGGGGCCCCG GGCCACGAGG CACGCGCCAG GCCCCCCTC	3600
	CGCTGATGGC GCTGTCCCTG ACCCCCCCGC ACGCGGACGG CCGCGCCCCG GTCCCGGAGC	3660
15	GAAAGCGGCC CTCTGCCGAC ACCATCGACC CCGCCGTTCCG GCGGGTTCTG CGATCCATAT	3720
	CCGAGCGCGC GCGGGTCGAG CGCATCAGCG AAAGCTTTGG ACGCAGTGCC CTGGTCATGC	3780
	AAGACCCCTT TGGCGGGATG CCGTTTCCCG CCGCGAACAG CCCCTGGGCT CCCGTGCTGG	3840
	CCACCCAAGC GGGGGGTTT GACGCCGAGA CCCGTCGGGT TTCCTGGGAA ACCCTGGTCG	3900
	CTCACGGCCC GAGCCTCTAC CGCACATTCG CAGCCAACCC GCGGGCCGCG TCGACAGCCA	3960
20	AGGCCATGCG CGACTGCGTG CTGCGCCAGG AAAATCTCAT CGAGGCCCTG GCGTCCGCGG	4020
	ATGAGACGCT GCGGTGGTGC AAGATGTGCA TTCACCACAA TCTGCCGCTC CGCCCCCAGG	4080
	ACCCATCAT CGGAACGGCG GCCGCCGTGC TGGAAAACCT CGCCACGCGC CTGCGCCCCCT	4140
	TTCTGCAGTG CTACCTGAAG GCCCGAGGCC TGTGCGGGCT GGACGACCTG TGCTCGCGGC	4200
	GACGCCTGTC GGACATTAAG GATATTGCCT CTTTGTGTGTT GGTTCATCCTG GCGCGCTCG	4260
25	CCAACCGCGT CGAGCGCGGC GTGTGCGAGA TCGACTACAC GACCGTGGGG GTTGGGGCCG	4320
	GCGAGACGAT GCACTTTTAC ATCCCGGGGG CTTGCATGGC GGTCTCATTT GAAATACTGG	4380
	ACACGCACCG CCAGGAGTGT TCCAGTCGCG TGTGCGAGCT GACGGCCAGT CACACTATCG	4440
	CCCCCTTATA TGTGCACGGC AAATACTTCT ACTGCAACTC CCTATTTTAG GCAAGAATAA	4500
	ACATATTGAC GTCAACCCAA GTGGTTCCGT GTGATGTTCT TGGCGCGCGC GCGGGTGGG	4560
30	GCGGAGACTC CGGGGCGATG CCGGCGTGCG CGTGGGAGGA GGGCGATGAC CCACCGGATA	4620
	AATGTGGGGC CCCGGCCCCG CCCGCTTCAT AGCGCGTCCA GGAACACG GCAGACGCGT	4680
	ATTACCGAC CCCCCCCTC GCAACATGAC AACGACGCC CTCTCGAACC TGTTTTTACG	4740
	GGCCCCGAC ATCACCCACG TCGCCCCCCC GTACTGTCTG AATGCCACGT GGCAGGCCGA	4800
	AAACGCCCTG CACACGACCA AAACGGACCC CGCGTGCCTG GCCGCGCGGA GTTATTTAGT	4860
35	CCGCGCCTCC TGCTCGACCA GCGGCCCCAT CCACTGTTTT TTCTTTGCGG TGTACAAGGA	4920
	CTCGCAGCAC TCCCTTCCGC TGGTTACCGA GTCGCCAAC TTCGCGGACC TGGTCAACCA	4980
	CCCCCCGTC TTGCGCGAAC TAGAGGATAA GCGTGGGGGG CGGCTGCGGT GCACGGGCCC	5040
	ATTAGCTGC GGAACCATCA AGGACGTCTC CGGTGCATCC CCCGCGGGG AATACACGAT	5100
	AAACGGTATC GTGTACCACT GTCACGTGCG GTATCCGTTT TCCAAAACCT GCTGGCTCGG	5160
40	GGCATCCGCG GCCCTACAAC ACCTTCGCTC TATAAGCTCA AGCGGCACGG CCGCTCGCGC	5220
	GGCAGAACAG CGACGCCACA AAATCAAAAT CAAAATCAAG GTATAACCCA CCCCCCTCC	5280
	TCCGAGTCCG TATGCAACCT CATTAATAAA GAGTGAGAAC CAACCAAAAC AGACGCGGTG	5340
	TGAGTTTGTG GGTATAGGA ACCCGGTAA TACCACGCGA CGAACCAGCG TGTGTGTTAA	5400

CGCGACTTTT ATTTCGTTGTA TCGCGGGAGG GGGGAAGCTT ACCGCCAAAG GAAGGCCAAG 5460  
 ATGATAACGA CGACCACCGC GACCACCCCA AAAACCGCAT GACGACACGT CCCGCCACAC 5520  
 CACCCTGGGG CTTGGGGCGT GTCGAGGCTC GACGCACAGC GGGCCGCGCG TTGGGCCCCG 5580  
 TACAGCTCTC GCGAATTGAC GAGCGGGGGT GCCCAGTGC GCGAGCTTTG CACGCGGGGT 5640  
 5 TGGTCGGCCG GCCCCACGGA CCCGCCCGGT GGCTCGGTGC GACATGCGGC CATGACCATG 5700  
 GCGTAGGTGG GGGGGCGATC CGAGGTGCGC TCTGCGTAAG TAGGGAGGCC CGACGGGAGG 5760  
 TCGCCTCCCA CGCCAGGGTG GGCCCAATC ATAGTTTCCG GTAGAAACAG GGGGTCTCC 5820  
 ACAAACAACC CCCCTGGGCC AAAGCTCCGG CGCCGCGCCC GTCGTTCCGC GCGGCGCCTG 5880  
 GCGCGCCGAG CGGCCCGCCA GCGGCGCGG CGCGAGCGGC CACGCTCACA CACCTCGCCG 5940  
 10 TCACCGGAAG AAGCCGTGA AACAGCCCA ACCGGCGACG TCCCTGCAGA GTACGGTGA 6000  
 GCGAGTCCG TGGGGGTGTC GATATCAATA ACGACAACT GGCCCGCGCT CGCGCCGGCC 6060  
 AACTCTCGT ATGGGGCGG GCGTCAATC ACGCTATCAT CTCCGTCATC CCTGCATGCG 6120  
 TGGGCATGCC CAGCCCCAA CGCATGGTG GGGATTGCG GCTCAGAAGC CTGCATGTCG 6180  
 TGTGGTCGGT CGTAGTCCAA CGTGCTCCC CCACCCACCA CACAGCCGGT CCCCACGCCG 6240  
 15 ACCACTAGAC CGCAGACGTC GCCCAACCGA GGTCCCCGTG CACAGACCGC GCCTTTTATA 6300  
 GCCCCAGGGG TTGCTAATTA ACGCACGCAT GCAGACGCAA TTTATTTTGC TCCCCGCGT 6360  
 CCTCCCTCC CCCGCTCCT CCCNT 6386

## (2) INFORMATION FOR SEQ ID NO:190:

20

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 477 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

25

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

30

Xaa Xaa Xaa Xaa Xaa Thr Arg Arg Ile Cys Arg Pro Ala Leu Pro Pro  
 1 5 10 15  
 Gly Gly Leu Ala Val Gly Gly Gln Met Tyr Val Asn Arg Asn Glu Ile  
 20 25 30  
 35 Phe Asn Ala Ala Val Thr Asn Ile Ile Leu Asp Leu Asp Ile Ala Leu  
 35 40 45  
 Lys Glu Pro Val Pro Phe Pro Arg Leu His Glu Ala Leu Gly His Phe  
 50 55 60  
 Arg Arg Gly Ala Ala Val Gln Leu Leu Phe Pro Ala Ala Arg Val Asp  
 40 65 70 75 80  
 Pro Asp Ala Tyr Pro Cys Tyr Phe Phe Lys Ser Ala Cys Arg Pro Arg  
 85 90 95  
 Ala Pro Pro Val Cys Ala Gly Asp Gly Pro Ser Ala Gly Gly Asp Asp

		100		105		110	
	Gly Asp	Gly Asp Trp Phe Pro Asp	Ala Gly Gly Asp Asp	Gly Asp Glu			
	115	120	125				
5	Glu Trp	Glu Glu Asp Thr Asp Pro Met	Asp Thr Thr His Gly Pro Leu				
	130	135	140				
	Pro Asp	Asp Glu Ala Ala Tyr Leu Asp	Leu Leu His Glu Gln Ile Pro				
	145	150	155	160			
	Ala Ala	Thr Pro Ser Glu Pro Asp Ser	Val Val Cys Ser Cys Ala Asp				
		165	170	175			
10	Lys Ile	Gly Leu Arg Val Cys Leu Pro	Val Pro Ala Pro Tyr Val Val				
		180	185	190			
	His Gly	Ser Leu Thr Met Arg Gly Val	Ala Arg Val Ile Gln Gln Ala				
		195	200	205			
	Val Leu	Leu Asp Arg Asp Phe Val Glu	Ala Val Gly Ser His Val Lys				
15		210	215	220			
	Asn Phe	Leu Leu Ile Asp Thr Gly Val	Tyr Ala His Gly His Ser Leu				
	225	230	235	240			
	Arg Leu	Pro Tyr Phe Ala Lys Ile Gly	Pro Asp Gly Ser Ala Cys Gly				
		245	250	255			
20	Arg Leu	Leu Pro Val Phe Val Ile Pro	Pro Ala Cys Glu Asp Val Pro				
		260	265	270			
	Ala Phe	Val Ala Ala His Ala Asp Pro	Arg Arg Phe His Phe His Ala				
		275	280	285			
	Pro Pro	Met Phe Ser Ala Ala Pro Arg	Glu Ile Arg Val Leu His Ser				
25		290	295	300			
	Leu Gly	Gly Asp Tyr Val Ser Phe Phe	Glu Lys Lys Ala Ser Arg Asn				
	305	310	315	320			
	Ala Leu	Glu His Phe Gly Arg Arg Glu	Thr Leu Thr Glu Val Leu Gly				
		325	330	335			
30	Arg Tyr	Asp Val Arg Pro Asp Ala Gly	Glu Thr Val Glu Gly Phe Ala				
		340	345	350			
	Ser Glu	Leu Leu Gly Arg Ile Val Ala	Cys Ile Glu Ala His Phe Pro				
		355	360	365			
	Glu His	Ala Arg Glu Tyr Gln Ala Val	Ser Val Arg Arg Ala Val Ile				
35		370	375	380			
	Lys Asp	Asp Trp Val Leu Leu Gln Leu	Ile Pro Gly Arg Gly Ala Leu				
	385	390	395	400			
	Asn Gln	Ser Leu Ser Cys Leu Arg Phe	Lys His Gly Arg Ala Ser Arg				
		405	410	415			
40	Ala Thr	Ala Arg Thr Phe Leu Ala Leu	Ser Val Gly Thr Asn Asn Arg				
		420	425	430			
	Leu Cys	Ala Ser Leu Cys Gln Gln Cys	Phe Ala Thr Lys Cys Asp Asn				
		435	440	445			

Asn Arg Leu His Thr Leu Phe Thr Val Asp Ala Gly Thr Pro Cys Ser  
 450 455 460  
 Arg Ser Ala Pro Ser Ser Thr Ser Arg Pro Ser Ser Ser  
 465 470 475

5

## (2) INFORMATION FOR SEQ ID NO:191:

## (i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 332 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

Met Leu Ala Val Arg Ser Leu Gln His Leu Thr Thr Val Ile Phe Ile  
 1 5 10 15  
 20 Thr Ala Tyr Gly Leu Val Leu Ala Trp Tyr Ile Val Phe Gly Asp Leu  
 20 25 30  
 His Arg Cys Ile Tyr Ala Val Arg Pro Ala Gly Ala His Asn Asp Thr  
 35 40 45  
 Ala Leu Val Trp Met Lys Ile Asn Gln Thr Leu Leu Phe Leu Gly Pro  
 25 50 55 60  
 Pro Thr Ala Pro Pro Gly Gly Ala Trp Thr Pro His Ala His Val Cys  
 65 70 75 80  
 Tyr Ala Asn Ile Ile Glu Gly Arg Ala Val Ser Leu Pro Ala Ile Pro  
 85 90 95  
 30 Gly Ala Met Ser Arg Arg Val Met Asn Val His Glu Ala Val Asn Cys  
 100 105 110  
 Leu Glu Ala Leu Trp Asp Thr Gln Met Arg Leu Val Val Val Gly Trp  
 115 120 125  
 Phe Leu Tyr Leu Ala Phe Val His Gln Arg Arg Cys Met Phe Gly Val  
 35 130 135 140  
 Val Ser Pro Ala His Ser Met Val Ala Pro Ala Thr Tyr Leu Leu Asn  
 145 150 155 160  
 Tyr Ala Gly Arg Ile Val Ser Ser Val Phe Leu Gln Tyr Pro Tyr Thr  
 165 170 175  
 40 Lys Ile Thr Arg Leu Leu Cys Glu Leu Ser Val Gln Arg Gln Thr Leu  
 180 185 190  
 Val Gln Leu Phe Glu Ala Asp Pro Val Thr Phe Leu Tyr His Arg Pro  
 195 200 205

Ala Val Gly Val Ile Val Gly Cys Glu Leu Leu Leu Arg Phe Val Gly  
 210 215 220  
 Leu Ile Val Gly Thr Ala Leu Ile Ser Arg Gly Ala Cys Ala Ile Thr  
 225 230 235 240  
 5 Tyr Pro Leu Phe Leu Thr Ile Thr Thr Trp Cys Phe Val Ser Ile Ile  
 245 250 255  
 Ala Leu Thr Glu Leu Tyr Phe Ile Leu Arg Arg Asp Ser Ala Pro Lys  
 260 265 270  
 Asn Ala Glu Pro Ala Ala Pro Arg Gly Arg Ser Lys Gly Trp Ser Gly  
 10 275 280 285  
 Val Cys Gly Arg Cys Cys Ser Ile Ile Leu Ser Gly Ile Ala Val Arg  
 290 295 300  
 Leu Cys Tyr Ile Ala Val Val Ala Gly Val Val Leu Met Ala Leu Arg  
 305 310 315 320  
 15 Tyr Glu Gln Glu Ile Gln Arg Arg Leu Phe Asp Leu  
 325 330

## (2) INFORMATION FOR SEQ ID NO:192:

## 20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 574 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

25

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:

30 Val Thr Pro Asp Gly Glu Gly Gln Gly Gly Val Ser Glu Ser Arg Pro  
 1 5 10 15  
 Arg Ser Cys Gly Tyr Lys Gly Ser His Arg Pro Thr Gly Arg Cys Val  
 20 25 30  
 Leu Pro Cys Ala Asp Pro Gly Cys Ala Ser Val Pro Leu Leu Asp Ser  
 35 35 40 45  
 Asp Pro Ala Thr Leu Phe Arg His Ala Pro Pro Arg Arg Thr Pro Ala  
 50 55 60  
 Ile Pro Ala Pro Ala Thr Tyr Asn Met Ala Thr Asp Ile Asp Met Leu  
 65 70 75 80  
 40 Ile Asp Leu Gly Leu Asp Leu Ser Asp Ser Glu Leu Glu Glu Asp Ala  
 85 90 95  
 Leu Glu Arg Asp Glu Glu Gly Arg Arg Asp Asp Pro Glu Ser Asp Ser  
 100 105 110

Ser Gly Glu Cys Ser Ser Ser Asp Glu Asp Met Glu Asp Pro Cys Gly  
 115 120 125  
 Asp Gly Gly Ala Glu Ala Ile Asp Ala Ala Ile Pro Lys Gly Pro Pro  
 130 135 140  
 5 Ala Arg Pro Glu Asp Ala Gly Thr Pro Glu Ala Ser Thr Pro Arg Pro  
 145 150 155 160  
 Ala Ala Arg Arg Gly Ala Asp Asp Pro Pro Pro Ala Thr Thr Gly Val  
 165 170 175  
 Trp Ser Arg Leu Gly Thr Arg Arg Ser Asp Arg Glu Pro His Gly Gly  
 10 180 185 190  
 Lys Val Ala Arg Ile Gln Pro Pro Ser Thr Lys Ala Pro His Pro Arg  
 195 200 205  
 Gly Gly Arg Arg Gly Arg Arg Arg Gly Arg Gly Arg Tyr Gly Pro Gly  
 210 215 220  
 15 Gly Ala Asp Ser Thr Pro Asn Pro Arg Arg Arg Val Ser Arg Asn Ala  
 225 230 235 240  
 His Asn Gln Gly Gly Arg His Pro Ala Ser Ala Arg Thr Asp Gly Pro  
 245 250 255  
 Gly Ala Thr His Gly Glu Ala Arg Arg Gly Gly Glu Gln Leu Asp Val  
 20 260 265 270  
 Ser Gly Gly Pro Arg Pro Arg Gly Thr Arg Gln Ala Pro Pro Pro Leu  
 275 280 285  
 Met Ala Leu Ser Leu Thr Pro Pro His Ala Asp Gly Arg Ala Pro Val  
 290 295 300  
 25 Pro Glu Arg Lys Ala Pro Ser Ala Asp Thr Ile Asp Pro Ala Val Arg  
 305 310 315 320  
 Ala Val Leu Arg Ser Ile Ser Ala Ala Val Glu Arg Ile Ser Glu Ser  
 325 330 335  
 Phe Gly Arg Ser Ala Leu Val Met Gln Asp Pro Phe Gly Gly Met Pro  
 30 340 345 350  
 Phe Pro Ala Ala Asn Ser Pro Trp Ala Pro Val Leu Ala Thr Gln Ala  
 355 360 365  
 Gly Gly Phe Asp Ala Glu Thr Arg Arg Val Ser Trp Glu Thr Leu Val  
 370 375 380  
 35 Ala His Gly Pro Ser Leu Tyr Arg Thr Phe Ala Ala Asn Pro Arg Ala  
 385 390 395 400  
 Ala Ser Thr Ala Lys Ala Met Arg Asp Cys Val Leu Arg Gln Glu Asn  
 405 410 415  
 Leu Ile Glu Ala Ser Ala Asp Glu Thr Leu Ala Trp Cys Lys Met Cys  
 40 420 425 430  
 Ile His His Asn Leu Pro Leu Arg Pro Gln Asp Pro Ile Ile Gly Thr  
 435 440 445  
 Ala Ala Ala Val Leu Glu Asn Leu Ala Thr Arg Leu Arg Pro Phe Leu  
 495

450                      455                      460  
 Gln Cys Tyr Leu Lys Arg Leu Cys Gly Leu Asp Asp Leu Cys Ser Arg  
 465                      470                      475                      480  
 Arg Arg Leu Ser Asp Ile Lys Asp Ile Ala Ser Phe Val Leu Val Ile  
 5                      485                      490                      495  
 Leu Ala Arg Leu Ala Asn Arg Val Glu Arg Gly Val Ser Glu Ile Asp  
                     500                      505                      510  
 Tyr Thr Thr Val Gly Val Gly Ala Gly Glu Thr Met His Phe Tyr Ile  
                     515                      520                      525  
 10    Pro Gly Ala Cys Met Ala Gly Leu Ile Glu Ile Leu Asp Thr Gln Glu  
                     530                      535                      540  
 Cys Ser Ser Arg Val Cys Glu Leu Thr Ala Ser His Thr Ile Ala Pro  
 545                      550                      555                      560  
 Leu Tyr Val His Gly Lys Tyr Phe Tyr Cys Asn Ser Leu Phe  
 15                      565                      570

## (2) INFORMATION FOR SEQ ID NO:193:

## (i) SEQUENCE CHARACTERISTICS:

- 20                      (A) LENGTH: 212 amino acids  
                     (B) TYPE: amino acid  
                     (C) STRANDEDNESS: single  
                     (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:193:

Met Trp Gly Pro Gly Pro Ala Arg Phe Ile Ala Arg Pro Gly Thr His  
 30    1                      5                      10                      15  
 Gly Arg Arg Val Phe Thr Asp Pro Pro Pro Arg Asn Met Thr Thr Thr  
                     20                      25                      30  
 Pro Leu Ser Asn Leu Phe Leu Arg Ala Pro Asp Ile Thr His Val Ala  
                     35                      40                      45  
 35    Pro Pro Tyr Cys Leu Asn Ala Thr Trp Gln Ala Glu Asn Ala Leu His  
                     50                      55                      60  
 Thr Thr Lys Thr Asp Pro Ala Cys Leu Ala Ala Arg Ser Tyr Leu Val  
 65                      70                      75                      80  
 Arg Ala Ser Cys Ser Thr Ser Gly Pro Ile His Cys Phe Phe Phe Ala  
 40                      85                      90                      95  
 Val Tyr Lys Asp Ser Gln His Ser Leu Pro Leu Val Thr Glu Leu Arg  
                     100                      105                      110  
 Asn Phe Ala Asp Leu Val Asn His Pro Pro Val Leu Arg Glu Leu Glu  
                     496

115                      120                      125  
 Asp Lys Arg Gly Gly Arg Leu Arg Cys Thr Gly Pro Phe Ser Cys Gly  
      130                      135                      140  
 Thr Ile Lys Asp Val Ser Gly Asp Ala Gly Glu Tyr Thr Ile Asn Gly  
 5    145                      150                      155                      160  
 Ile Val Tyr His Cys His Cys Arg Tyr Pro Phe Ser Lys Thr Cys Trp  
                          165                      170                      175  
 Leu Gly Ala Ser Ala Ala Leu Gln His Leu Arg Ser Ile Ser Ser Ser  
                          180                      185                      190  
 10 Gly Thr Ala Ala Arg Ala Ala Glu Gln Arg Arg His Lys Ile Lys Ile  
                          195                      200                      205  
 Lys Ile Lys Val  
      210

15                      (2) INFORMATION FOR SEQ ID NO:194:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 117 amino acids  
 (B) TYPE: amino acid  
 20 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

25                      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

Met Ile Gly Ala His Pro Gly Val Gly Gly Asp Leu Pro Ser Gly Leu  
      1                      5                      10                      15  
 Pro Thr Tyr Ala Glu Ala Thr Ser Asp Arg Pro Pro Thr Tyr Ala Met  
 30                      20                      25                      30  
 Val Met Ala Ala Cys Pro Thr Glu Pro Pro Gly Gly Ser Val Gly Pro  
                          35                      40                      45  
 Ala Asp Gln Pro Arg Val Gln Ser Ser Arg Thr Trp Arg Pro Pro Leu  
                          50                      55                      60  
 35 Val Asn Ser Arg Glu Leu Tyr Arg Ala Gln Arg Ala Ala Arg Cys Ala  
      65                      70                      75                      80  
 Ser Ser Ser Asp Thr Pro Gln Ala Pro Gly Trp Cys Gly Gly Thr Cys  
                          85                      90                      95  
 Arg His Ala Val Phe Gly Val Val Ala Val Val Val Val Ile Ile Leu  
 40                      100                      105                      110  
 Ala Phe Leu Trp Arg  
      115



## (2) INFORMATION FOR SEQ ID NO:195:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 3699 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

GGGGGCCACG CGGCGGCGGG CCTGACGGAG CTGTGTCAGA CCCTCGCGCC CCGGGACCTC 60  
 ACGGACCCGC TGCTGTTTGC GTACGTCGGA TTCCAGGTCG TGAACCACGG GCTGATGTTT 120  
 GTGGTCCCCG ACATCGCCGT ATACGCGATG CTGGGGGGCG CCGTGTGGAT CTCGCTGACG 180  
 15 CAGGTGCTTG GGCTCCGGCG CCGCCTTCAC AAGGACCCAG ACGCCGGGCC CTGGGCGGCC 240  
 GCGACCCTGC GGGGCCTCTT TTTCTCCGTC TACGCATTGG GGTTTGCGGC GGGGGTGCTG 300  
 GTGCGGCCGC GGATGGCGGC GAGCCGGCGG TCGGGGTGAT CGCCATTTCA AATAAAAGGC 360  
 ACGAGTTCCT CGAATACCAC CGGCGTGTGA TGATTTCGCC CTACCGCTCC GATCCCCGGG 420  
 GGGAGGGGGG AAGGAAATGG GGGCGGGGGT GCCGTGGACG GGTATAAAGG CCAGGGGGGC 480  
 20 AGGCGGGCCC ATCACTGTTA GGGTGTTAGG TTGGGAGGTG GCACAAAAG CGACACACCC 540  
 GTGTTGTAGT TGTCGCGGG AGGCGGTGGT TTCCGGCAAC CCTCCTCGCT GCGCCGGGCG 600  
 CGCCACCCG TCCTTCGCGG GGGCCGGGGC TCTTCTGGTC ATGGCCCTTG GACGGGTGGG 660  
 CCTAGCCGTG GGCTGTGGG GCCTGCTGTG GGTGGGTGTG GTCGTGGTGC TGGCCAATGC 720  
 CTCCCCGGA CGCACGATAA CGGTGGGCCC GCGGGGAAC GCGAGCAATG CCGCCCCCTC 780  
 25 CGCGTCCCCG CGGAACGCAT CCGCCCCCG AACCACACCC ACGCCCCCC AACCCGCAA 840  
 GGCGACGAAA AGTAAGGCCT CCACCGCAA ACCGGCCCC CCCCCAAGA CCGGGCCCC 900  
 GAAGACATCC TCGGAGCCCG TCGATGCAA CCGCCACGAC CCGCTGGCCC GGTACGGCTC 960  
 GCGGTGCAA ATCCGATGCC GGTTCCTCAA CTCCACCCGC ACGGAGTCCC GCCTCCAGAT 1020  
 CTGGCGTTAT GCCACGGCGA CGGACGCCGA GATCGGAACG GCGCCTAGCT TAGAGGAGGT 1080  
 30 GATGGTAAAC GTGTCGGCCC CGCCGGGGG CCAACTGGTG TATGACAGCG CCCCCAACCG 1140  
 AACGGACCCG CACGTGATCT GGGCGGAGGG CGCCGGCCCC GCGCCAGCC CGCGGCTGTA 1200  
 CTCGGTCGTC GGGCCGCTGG GTCGGCAGCG GCTCATCATC GAAGAGCTGA CCCTGGAGAC 1260  
 CCAGGGCATG TACTACTGGG TGTGGGGCCG GACGGACCGC CCGTCCGCGT ACGGGACCTG 1320  
 GGTGCGCGTT CGCGTGTTC GCCCTCCGTC GCTGACCATC CACCCCCACG CCGTGTCTGGA 1380  
 35 GGGCCAGCCG TTTAAGGCGA CGTGCACGGC CGCCACCTAC TACCCGGGCA ACCGCGCGGA 1440  
 GTTCGTCTGG TTCGAGGACG GTCGCCGGGT ATTCGATCCG GCCCAGATAC ACACGCAGAC 1500  
 GCAGGAGAAC CCCGACGGCT TTTCCACCGT CTCCACCGTG ACCTCCGCGG CCGTCGGCGG 1560  
 CCAGGGCCCC CCGCGCACCT TCACCTGCCA GCTGACGTGG CACCGCGACT CCGTGTCTGT 1620  
 CTCTCGGCGC AACGCCAGCG GCACGGCATC GGTGCTGCCG CGGCCAACCA TTACCATGGA 1680  
 40 GTTTACGGGC GACCATGCGG TCTGCACGGC CGGCTGTGTG CCCGAGGGGG TGACGTTTGC 1740  
 CTGGTTCTTG GGGGACGACT CCTCGCCGGC GGAGAAGGTG GCCGTCGCGT CCCAGACATC 1800  
 GTGCGGGCGC CCCGGCACCG CCACGATCCG CTCCACCCTG CCGGTCTCGT ACGAGCAGAC 1860  
 CGAGTACATC TGCCGGCTGG CGGGATACCC GGACGGAATT CCGGTCTTAG AGCACCACGG 1920

	CAGCCACCAG	CCCCGCCGC	GGGACCCAC	CGAGCGGCAG	GTGATCCGGG	CGGTGGAGGG	1980
	GGCGGGGATC	GGAGTGGCTG	TCCTTGTCGC	GGTGGTTCTG	GCCGGGACCG	CGGTAGTGTA	2040
	CCTCACCAC	GCCTCCTCGG	TGCGCTATCG	TCGGCTGCGG	TAACGCCGG	GCCGGGCCCC	2100
	GCCGCCGTT	GTCTTCTTTT	CCACCCCTTC	CGTCCCCCGT	ACCCACCACA	CCCCACCCCA	2160
5	CCCCCCCC	GTCCCCGGG	CGTTATAAGC	CGCCGCACTC	GCTTTTCCCA	CCGGAATC	2220
	CTCGGCCCGA	TCCGAACGGC	GCACGCCGCG	TGGGCTCCAA	ACGCCTCCGG	AAGAGAGCGC	2280
	CCCGCCCCGA	TATTCAAGCC	CGCGGTGGTG	CTATGGCTTT	CCGTGCTTCG	GGACCCGCCT	2340
	ACCAGCCCT	CGCCCCGCG	GCCTCCCCGG	CGCGGGCTCG	TGTTCGGGCC	GTGGCCTGGA	2400
	TCGGCGTCGG	AGCGATCGTC	GGGGCCTTTG	CGCTCGTCGC	CGCGTTGGTT	CTCGTACCCC	2460
10	CTCGGTCTC	GTGGGGACTC	TCGCCGTGCG	ACAGCGGCTG	GCAGGAATTC	AACCGGGGAT	2520
	GCGTCGCGTG	GGACCCACC	CCCGTCGAGC	ACGAGCAGGC	GGTCGGCGGC	TGCAGCGCGC	2580
	CGGCCACCCT	TATCCCCCGT	GCGGCCGCCA	AGCACCTGGC	CGCTCTGACA	CGCGTCCAGG	2640
	CGGAGAGATC	GTCCGGTTAC	TGGTGGGTGA	ACGGAGACGG	CATCCGGACC	TGTCTGAGAC	2700
	TCGTCGACAG	CGTCAGTGGC	ATCGACGAGT	TTTGCGAGGA	GCTCGCGATC	CGCATATGCT	2760
15	ACTACCACG	AAGCCCCGGC	GGGTTTGTCC	GCTTCGTAAC	TTCGATACGT	AACGCCCTGG	2820
	GGTTGCCGTG	AGGCGCGCGT	CCGACGGTCC	CGCTTCTCGC	CTCTCTTCTT	CCCCCTCCCC	2880
	ACCCACCCCA	CCGACCAACG	ACGGCGTTTG	GCCAATACCC	TCCTTTTTTC	TTTTTCTCTT	2940
	CCCCCCCCAA	AAAAAAAAC	AATAAACAGC	TAATTGCGTA	CGACAAACCA	TGCGGAATC	3000
	GCTGTTTTTT	TTTCTCTGTT	TGTTACTTTT	TATTGAAAAC	AGACATACGG	GGAAAGGGGC	3060
20	CGGAAACCGA	GACGGTGGGG	CCGGCGGTTCG	CATTTTTTTA	ATGGCTCTGG	TGTCGGCCGC	3120
	GTTTGAGCTT	CGTCAACAGG	GCGCTGAGGG	CGGCGACGTT	TGTCGGGCCG	TCGTTGGCCA	3180
	GCGCGTTGGT	CCGGGGGCGG	GCGGGCATGG	GCGACAGGCT	TAGTCCCGGG	TCCGGGGCGC	3240
	GTGTGGCCCC	CGGAGGGGAG	AAGAGGGCAG	ACCCGCCCCA	GTCGTACAGG	GGATTTTCCG	3300
	CCTCGATGTA	CGGGGAGTCC	GGGGCGTCTC	CCGGCGGGGC	CGCCCCGCCG	GCGTCTTGCC	3360
25	GGCGAAGGCA	GATGTTTTTCG	TATACCCGAA	CCCAGGGGAT	CTCCTCGTAG	ACGCGCCCCC	3420
	CATCCTCGCT	CACCGACTCG	TAAATGGAAT	CTGCGTCCTC	GGAGGGGGCG	CGGGGGGCGT	3480
	GGCTTTCGGC	CGGCCAGGCG	GCGGCGGTGG	TGTCGGCGGC	GGGGGTGGCG	CCAAGCCCGA	3540
	CGCCCGCGGG	CATGGCGGCG	TCATCGTCGG	GCAGCAGATA	CGTGTTTTCC	ATCTGGTCCG	3600
	GTTCCGGCTC	CGCGTCTGGC	CCCCAGGTCC	GCACCGCGTC	GTAAACCCCG	GCGGCCTCGC	3660
30	GCTGAGCCGC	GAGCGGGCGC	GCCGCGGCTG	CCGGCCGC			3699

## (2) INFORMATION FOR SEQ ID NO:196:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 117 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 40 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

Xaa Xaa Xaa Xaa Xaa Gly Gly His Ala Ala Ala Gly Leu Thr Glu Leu  
 1 5 10 15  
 Cys Gln Thr Leu Ala Pro Arg Asp Leu Thr Asp Pro Leu Leu Phe Ala  
 20 25 30  
 5 Tyr Val Gly Phe Gln Val Val Asn His Gly Leu Met Phe Val Val Pro  
 35 40 45  
 Asp Ile Ala Val Tyr Ala Met Leu Gly Gly Ala Val Trp Ile Ser Leu  
 50 55 60  
 Thr Gln Val Leu Gly Leu Arg Arg Arg Leu His Lys Asp Pro Asp Ala  
 10 65 70 75 80  
 Gly Pro Trp Ala Ala Ala Thr Leu Arg Gly Leu Phe Phe Ser Val Tyr  
 85 90 95  
 Ala Leu Gly Phe Ala Ala Gly Val Leu Val Arg Pro Arg Met Ala Ala  
 100 105 110  
 15 Ser Arg Arg Ser Gly  
 115

## (2) INFORMATION FOR SEQ ID NO:197:

## 20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 536 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

25

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

30 Met Gly Ala Gly Val Pro Trp Thr Gly Ile Lys Arg Ala Gly Gly Pro  
 1 5 10 15  
 Ile Thr Val Arg Val Leu Gly Trp Glu Val Ala Gln Lys Ala Thr His  
 20 25 30  
 Pro Cys Cys Ser Cys Pro Arg Glu Ala Val Val Ser Gly Asn Pro Pro  
 35 35 40 45  
 Arg Cys Ala Gly Arg Ala His Arg Ser Phe Ala Gly Ala Gly Ala Leu  
 50 55 60  
 Leu Val Met Ala Leu Gly Arg Val Gly Leu Ala Val Gly Leu Trp Gly  
 65 70 75 80  
 40 Leu Leu Trp Val Gly Val Val Val Val Leu Ala Asn Asp Gly Arg Thr  
 85 90 95  
 Ile Thr Val Gly Pro Arg Gly Asn Asn Ala Ala Pro Ser Asp Arg Asn  
 100 105 110  
 500

Ala Ser Ala Pro Arg Thr Thr Pro Thr Pro Pro Gln Pro Arg Lys Ala  
 115 120 125  
 Thr Lys Ser Lys Ala Ser Thr Ala Lys Pro Ala Pro Pro Pro Lys Thr  
 130 135 140  
 5 Gly Pro Pro Lys Thr Ser Ser Glu Pro Val Arg Cys Asn Arg His Asp  
 145 150 155 160  
 Pro Leu Ala Arg Tyr Gly Ser Arg Val Gln Ile Arg Cys Arg Phe Pro  
 165 170 175  
 Asn Ser Thr Arg Thr Glu Ser Arg Leu Gln Ile Trp Arg Tyr Ala Thr  
 10 180 185 190  
 Ala Thr Asp Ala Glu Ile Gly Thr Ala Pro Ser Leu Glu Glu Val Met  
 195 200 205  
 Val Asn Val Ser Ala Pro Pro Gly Gly Gln Leu Val Tyr Asp Ser Ala  
 210 215 220  
 15 Pro Asn Arg Thr Asp Pro His Val Ile Trp Ala Glu Gly Ala Gly Pro  
 225 230 235 240  
 Gly Asp Arg Lys Val Val Gly Pro Leu Gly Arg Gln Arg Leu Ile Ile  
 245 250 255  
 Glu Glu Leu Thr Leu Glu Thr Gln Gly Met Tyr Tyr Trp Val Trp Gly  
 20 260 265 270  
 Arg Thr Asp Arg Pro Ser Ala Tyr Gly Thr Trp Val Arg Val Arg Val  
 275 280 285  
 Phe Arg Pro Pro Ser Leu Thr Ile His Pro His Ala Val Leu Glu Gly  
 290 295 300  
 25 Gln Pro Phe Lys Ala Thr Cys Thr Ala Ala Thr Tyr Tyr Pro Gly Asn  
 305 310 315 320  
 Arg Ala Glu Phe Val Trp Phe Glu Asp Gly Arg Arg Val Phe Asp Pro  
 325 330 335  
 Ala Gln Ile His Thr Gln Thr Gln Glu Asn Pro Asp Gly Phe Ser Thr  
 30 340 345 350  
 Val Ser Thr Val Thr Ser Ala Ala Val Gly Gly Gln Gly Pro Pro Arg  
 355 360 365  
 Thr Phe Thr Cys Gln Leu Thr Trp His Arg Asp Ser Val Ser Phe Ser  
 370 375 380  
 35 Arg Arg Asn Ala Ser Gly Thr Ala Ser Val Leu Pro Arg Pro Thr Ile  
 385 390 395 400  
 Thr Met Glu Phe Thr Gly Asp His Ala Val Cys Thr Ala Gly Cys Val  
 405 410 415  
 Pro Glu Gly Val Thr Phe Ala Trp Phe Leu Gly Asp Asp Ser Ser Pro  
 40 420 425 430  
 Ala Glu Lys Val Ala Val Ala Ser Gln Thr Ser Cys Gly Arg Pro Gly  
 435 440 445  
 Thr Ala Thr Ile Arg Ser Thr Leu Pro Val Ser Tyr Glu Gln Thr Glu

```

      450              455              460
Tyr Ile Cys Arg Leu Ala Gly Tyr Pro Asp Gly Ile Pro Val Leu Glu
465              470              475              480
His His Gly Ser His Gln Pro Pro Pro Arg Asp Pro Thr Glu Arg Gln
5              485              490              495
Val Ile Arg Ala Val Glu Gly Ala Gly Ile Gly Val Ala Val Leu Val
      500              505              510
Ala Val Val Leu Ala Gly Thr Ala Val Val Tyr Leu Thr His Ala Ser
      515              520              525
10 Ser Val Arg Tyr Arg Arg Leu Arg
      530              535

```

## (2) INFORMATION FOR SEQ ID NO:198:

## 15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 189 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

```

25 Val Gly Ser Lys Arg Leu Arg Lys Arg Ala Pro Arg Pro Asp Ile Gln
   1              5              10              15
Arg Gly Ala Met Ala Phe Arg Ala Ser Gly Pro Ala Tyr Gln Pro Leu
      20              25              30
Ala Pro Ala Asp Ala Arg Ala Arg Val Pro Ala Val Ala Trp Ile Gly
30      35              40              45
Val Gly Ala Ile Val Gly Ala Phe Ala Leu Val Ala Ala Leu Val Leu
   50              55              60
Val Pro Pro Arg Ser Ser Trp Gly Leu Ser Pro Cys Asp Ser Gly Trp
65              70              75              80
35 Gln Glu Phe Asn Ala Gly Cys Val Ala Trp Asp Pro Thr Pro Val Glu
      85              90              95
His Glu Gln Ala Val Gly Gly Cys Ser Ala Pro Ala Thr Leu Ile Pro
      100              105              110
Arg Ala Ala Ala Lys His Leu Ala Ala Leu Thr Arg Val Gln Ala Glu
40      115              120              125
Arg Ser Ser Gly Tyr Trp Trp Val Asn Gly Asp Gly Ile Arg Thr Cys
      130              135              140
Leu Arg Leu Val Asp Ser Val Ser Gly Ile Asp Glu Phe Cys Glu Glu

```

145                      150                      155                      160  
 Leu Ala Ile Arg Ile Cys Tyr Tyr Pro Arg Ser Pro Gly Gly Phe Val  
                                  165                      170                      175  
 Arg Phe Val Thr Ser Ile Arg Asn Ala Leu Gly Leu Pro  
 5                      180                      185

## (2) INFORMATION FOR SEQ ID NO:199:

## (i) SEQUENCE CHARACTERISTICS:

- 10                      (A) LENGTH: 198 amino acids  
                          (B) TYPE: amino acid  
                          (C) STRANDEDNESS: single  
                          (D) TOPOLOGY: linear

- 15                      (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

Gln Arg Pro Ala Ala Ala Arg Pro Leu Ala Ala Gln Arg Glu Ala  
 20                      1                      5                      10                      15  
 Ala Gly Val Tyr Asp Ala Val Arg Thr Trp Gly Pro Asp Ala Glu Ala  
                                  20                      25                      30  
 Glu Pro Asp Gln Met Glu Asn Thr Tyr Leu Leu Pro Asp Asp Ala  
                                  35                      40                      45  
 25                      Ala Met Pro Ala Gly Val Gly Leu Gly Ala Thr Pro Ala Ala Asp Thr  
                                  50                      55                      60  
 Thr Ala Ala Ala Trp Pro Ala Glu Ser His Ala Pro Arg Ala Pro Ser  
 65                      70                      75                      80  
 Glu Asp Ala Asp Ser Ile Tyr Glu Ser Val Ser Glu Asp Gly Gly Arg  
 30                      85                      90                      95  
 Val Tyr Glu Glu Ile Pro Trp Val Arg Val Tyr Glu Asn Ile Cys Leu  
                                  100                      105                      110  
 Arg Arg Gln Asp Ala Gly Gly Ala Ala Pro Pro Gly Asp Ala Pro Asp  
                                  115                      120                      125  
 35                      Ser Pro Tyr Ile Glu Ala Glu Asn Pro Leu Tyr Asp Trp Gly Gly Ser  
                                  130                      135                      140  
 Ala Leu Phe Ser Pro Pro Gly Ala Thr Arg Ala Pro Asp Pro Gly Leu  
 145                      150                      155                      160  
 Ser Leu Ser Pro Met Pro Ala Arg Pro Arg Thr Asn Ala Asn Asp Gly  
 40                      165                      170                      175  
 Pro Thr Asn Val Ala Ala Leu Ser Ala Leu Leu Thr Lys Leu Lys Arg  
                                  180                      185                      190  
 Gly Arg His Gln Ser His

195

## (2) INFORMATION FOR SEQ ID NO:200:

## 5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 152 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

```
CTGTGTGAAA TTGTTATCCG CTCACAATTC CACACAACAT ACGAGCCGGA AGCATAAAGT      60
15 GTAAAGCCTG GGGTGCCTAA TGAGTGAGCT AACTCACATT AATTGCGTTG CGCTCACTGC      120
CCGCTTTCCA GTCGGGAAAC CTGTCGTGCC A                                     152
```

## (2) INFORMATION FOR SEQ ID NO:201:

## 20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 129 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

```
CCCGGAGCCC GGCGGCCGCA GCCGAGCAGC GCCGCGGGCT CCGGGGCCGG GCCGGGCCGG      60
30 CAACGCCCCG CGCCGGCCGC GCGGGTGAGA ACCCCTGTGT CATTGTTTAC GTGGCCCGCG      120
GCCAGCAG                                     129
```

## (2) INFORMATION FOR SEQ ID NO:202:

## 35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 127 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

GGTGGGGCGC GCGGGCCGGC TCGGGGTGGG GGGAGAGTGT CGTGGGTGTG TTTTCGTGTC 60  
 CCCCACCACC ACTCCCACCC CGACCGCCGC CGCGCCCGCG TTTCTGCCGC CCGCGCGCTC 120  
 CTGTGT 127

5 (2) INFORMATION FOR SEQ ID NO:203:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 157 base pairs  
 (B) TYPE: nucleic acid  
 10 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

15 ATGCACGTGT AACCGCCAGT CCGTGCTTGC CTAGCGAACT CACCCGTCCC GGCTGGCGTG 60  
 CGCAGCCCGG GCCGTGTTGC GGGCCCTCTT AAGGGGCGGC GGCAGGACGG GGA CTCCGCC 120  
 CCGCCTCCTT TCCCCGCGG AGTCAACCCC CGGGGG 157

20 (2) INFORMATION FOR SEQ ID NO:204:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16813 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:

30 GGGGGGACGG GACGGGGGGA CGGGACGGGG GGGCCCCGAT CCCAACATCC GCGCTTTCTC 60  
 GCAGGCCGGG CGCCGCCCTC GTGGACGGGA CACCGGTGTG GTA ACTGGCG ACAAGGCGTC 120  
 GCCACTATGG CAGACATCCC CCCGGACCCG CCCGCGCTCA ACACGACGCC TGCGAATCAT 180  
 GCTCCCCCAT CCCCACCCCC GGGTTCACGG AAGCGCAGAC GCCCCGTCCT CCCCAGCTCG 240  
 35 TCGGAATCTG AGGGTAAGCC CGACACAGAA TCGGAATCCT CCTCGACCGA GTCGTCCGAG 300  
 GATGAGGCGG GAGACCTACG CGGCGGGCGC CGTCGCTCCC CGCGGGAGCT CGGGGGGAGG 360  
 TATTTTTTGG ATCTGTGCGC AGAATCGACC ACGGGGACGG AATCGGAGGG AACGGGGCCG 420  
 TCGGACGACG ATGATGATGA TGCGTCAGAC GGCTGGTTGG TTGACACCCC CCCCCGAAA 480  
 TCCAAGCGAC CCCGAATCAA CCTGCGATTA ACGAGCTCCC CCGACCGGCG CGCGGGTGTG 540  
 40 GTTTTCCCCG AGGTGTGGAG AAACGACAGA CCTATCCGCG CGGCGCAACC CCAGGCCCCG 600  
 GCCCAGTCTT CCGGGGATCG CGCAGCCGCA CCGCGGCGCT CTGCTCGCCA GGCCAGATG 660  
 CGGAGCGGAG CCGCCTGGAC GCTTGATCTG CATTACATAC GCCAGTGCGT CAACCAGCTC 720  
 TTTCGGATCC TGCGTGCCGC CCCGAACCCG CCCGGCAGCG CCAACCGCCT GCGCCACCTG 780



	GTGCGAGACT	GCTACCTTAT	GGGCTACTGC	CGGACCCGCC	TGGGGCCGCG	CACGTGGGGC	840
	CGCCTGCTGC	AGATCTCGGG	CGGAACCTGG	GACGTGCGCC	TGCGAAACGC	AATCCGGGAG	900
	GTCGAGGCGC	GTTTGAAC	CGCCGCCGAG	CCCGTGTGCG	AGCTGCCCTG	TCTGAACGCC	960
	AGGCGTTACG	GCCCCGAGTG	TGATGTTGGC	AATCTCGAGA	CCAACGGCGG	CTCGACGAGC	1020
5	GATGATGAGA	TATCGGATGC	GACGGACTCG	GACGATACCC	TCGCGTCCCA	TTCCGACACG	1080
	GAGGGGGGGC	CCTCCCCGGC	CGGCCGGGAG	AACCCGGAAT	CCGCGTCCGG	CGGGGCTATC	1140
	GCGGCTCGGC	TGGAGTGTGA	GTTTGGGACG	TTTGACTGGA	CGTCCGAGGA	GGGCTCCAG	1200
	CCCTGGCTGT	CCGCGGTGGT	CGCCGATACC	AGCTCCGCCG	AACGCTCTGG	CCTACCCGCC	1260
	CCGGGCGCGT	GTCGCGCAAC	GGAAGCCCCA	GAACGCGAGG	ACGGGTGCCG	AAAAATGCGC	1320
10	TTCCCCCGCG	CCTGCCCTTA	TCCCTGCGGC	CACACATTTT	TCCGGCCATG	AGCGCGGGAC	1380
	CCCCAGCCCG	GTGTGTTTGC	CAAACGAAAA	TAAACGCCCT	ACAAGAAAGC	TTTGTGTCT	1440
	GAGTGTCTGG	TTTTTCTGGG	GGTGGAGGAA	GGAACGACAA	AAAAAGAAAC	AAACGCGACA	1500
	CCGCTCGTAC	GTGTAATGGG	GCGCAGTGTT	TTTTATTAGC	ATCGGGGGGG	GGTTAGAGGT	1560
	TGGTGATTGG	ATAGCAAACG	TGGGATGACG	GAGGCCACTC	GTCGCCAACG	GCCAGCGGGG	1620
15	GCCCGGGGTT	CTGGGGGTCA	TCGTCCCCCG	TCTGCCAGGA	GGGCTCATCG	GGAATCTCGG	1680
	GTCGCCCCAT	GCACGTAAAA	CACGGGCGCT	GCGTGGGGTG	GGTCGCCGGA	TGCGGGCGGG	1740
	ATGATGCGGG	GCGGGGTTTG	TTGTGAGGAG	CCACGAGGGA	CCGTAGCCAG	CGAAGACAGC	1800
	TGCGTTCCCG	GTCGCCGGGC	ACCACCACGC	CGTATTGGTA	TTCGTATCGG	CTAAGGAGAT	1860
	TTTCCAGGGG	GTGATTAGGC	GCTGCGGGGA	ACGGGGTCCA	CGACACGGTC	CGCTCGGGCA	1920
20	AAAACCGATC	GGGCAGGGGC	CACGGTTCCC	CCACCCACGC	GTCGTGGGTC	TTCATGGCGA	1980
	TGAAGCGAAA	CCCCAGCCGG	GTTTTTTGTG	CGTACTCTAA	AAACGGCACA	CACAGGTCCG	2040
	CCGCCCCGAC	CACCCACAGG	TGGTATAGCC	GGTGGGGGCC	GGGGCGCTCT	TGATGCAGGA	2100
	GCCGAAAAACA	CGCAGGGGCA	TCCAGAATCT	CGATGCTTTC	CAGGGGGTCG	TCCTCCGCAA	2160
	ACAGGCCCGT	CGTGGTGTTC	GGGGGACAGC	GACAGGAGCG	GGTTCGCACG	ATCGGTGCGG	2220
25	TGAATTTGGG	CAAGTCCATC	AGAGGCTCGG	CCAGCCTGCG	AAGGTTGCC	GGGCGAACCA	2280
	CCACCGGGGT	TCCCAGAGGC	TCGGAGGCCA	GGATCCGGCA	TTGCCGAAGC	AGAAAACTCC	2340
	ACAGAGCCCG	GCTTTCGTCA	GCGGAAGTCC	GCGGCAGGGC	GTTTCGTTGG	TCTAGGAGGG	2400
	TAACCACACT	TACAACAACA	ACGCCCATGT	CGGTATATTA	GGCCCGTGGT	CCGATCTTCA	2460
	CTCACTCGCC	TGTCTGCGGA	CCTATGCACG	GCGGGACGGC	GCGCGGACCC	GGGGGGGCTG	2520
30	CTTGCTATCA	CACGCCCCGT	TCGCACGTTT	GATTTTTTCA	GCCTTGTTTG	GTTGGCTAGG	2580
	TATCCCGGAT	AATCTGACGT	TCCGGATATA	GGGGGCGGGG	GTAGTGGGGG	GGTGTGTGCA	2640
	CAAAC TGCCG	CTTCTTAAAA	CACCGGGGCC	CGTCGCTCGG	GGTGCTCGTT	GGTTGGCACG	2700
	CGCGACGCGG	CGAATGGCCT	GTCGTAAGTT	CTGTGGGGTC	TACCGTAGAC	CCGACAAGAG	2760
	ACAGGAGGCG	TCCGTCCCCG	CGGAGACAAA	CACGGCCCCG	GCCTTCCCCG	CGAGCACCTT	2820
35	TTATACCCCC	GCGGAGGATG	CGTACCTGGC	CCCCGGGGCC	CCGGAAACCA	TCCACCCTTC	2880
	CCGCCCCACG	TCCCCCGGCG	AGGCTGCGCG	CCTGTGTCAG	CTGCAGGAGA	TCTTGGCCCA	2940
	GATGCACAGC	GACGAGGACT	ACCCCATCGT	GGACGCCGCG	GGTGC GGAGG	AGGAAGACGA	3000
	GGCCGACGAT	GACGCCCCGG	ATGACGTGGC	CTACCCGGAG	GACTACGCGG	AGGGGCGTTT	3060
	TCTGTCCATG	GTTTCGGCCG	CCCCCTGCC	CGGAGCCAGC	GGCCATCCTC	CTGTTCCGGG	3120
40	CCGCGCAGCC	CCCCCGACG	TCCGGACCTG	CGACAGCGGT	AAGGTGGGGG	CCACGGGGTT	3180
	CACCCCGGAA	GAGCTCGACA	CCATGGACCG	GGAGGCACTT	CGGGCCATCA	GCCGCGGGTG	3240
	CAAGCCCCCT	TCGACCCTGG	CAAACTGGT	GACCGGGCTG	GGATTGCGGA	TCCACGGAGC	3300
	GCTCATCCCC	GGGTCGGAGG	GGTGTGTCTT	TGATAGCAGC	CACCCGAAC	ACCCTCATCG	3360

	GGTAATCGTC	AAGGCGGGGT	GGTACGCCAG	CACGAACCAC	GAGGCGCGGC	TGCTGAGACG	3420
	CCTGAACCAC	CCCGCGATCC	TACCCCTCCT	GGACCTGCAC	GTCGTTTCTG	GGGTCACGTG	3480
	TCTGGTCCTC	CCCAAGTATC	ACTGCGACCT	GTATACCTAT	CTGAGCAAGC	GCCCGTCTCC	3540
	GTTGGGCCAC	CTACAGATAA	CCGCGGTCTC	CCGGCAGCTC	TTGAGCGCCA	TCGACTACGT	3600
5	CCACTGCGAA	GGCATCATCC	ACCGCGATAT	TAAGACCGAG	AACATCCTCA	TCAACACCCC	3660
	CGAGAACATC	TGTCTGGGGG	ACTTTGGGGC	GGCGTGCTTT	GTGCGCGGGT	GTCGATCGAG	3720
	CCCTTTCAT	TACGGGATCG	CAGGCACCAT	CGATACAAAC	GCCCCCGAGG	TCCTGGCCCG	3780
	GGATCCGTAC	ACCCAGGTAA	TCGACATCTG	GAGCGCCGGC	CTGGTGATCT	TTGAGACCGC	3840
	CGTCCACACC	GCGTCCTTGT	TCTCGGCCCC	GCGCGACCCC	GAAAGGCGGC	CGTGCGACAA	3900
10	CCAGATCGCG	CGCATCATCC	GACAGGCCCA	GGTACACGTC	GACGAGTTTC	CAACGCACGC	3960
	GGAATCGCGC	CTCACC CGC	ACTACC GCTC	GCGGGCGGCC	GGGAACAATC	GTCCGGCGTG	4020
	GACCCGACCG	GCATGGACCC	GCTACTACAA	GATCCACACA	GACGTCGAAT	ATCTCATCTG	4080
	CAAAGCCCTT	ACCTTTGACG	CGGCGCTCCG	CCCAAGCGCC	GCGGAGTTGC	TGCGCCTGCC	4140
	GCTATTTTAC	CCTAAGTGAC	CCCGCTCCCC	CCGGGGGGCG	TGGAGGGGGG	GCTGGTTGGA	4200
15	TGTTTTTGCA	CAAAAAGACG	CGGCCCTCGG	GCTTTGGTGT	TTTTGGCACC	TTGCCGCCCC	4260
	GCGTCATGCA	CGCCATCGCT	CCCAGGTTGC	TTCTTCTTTT	TGTTCTTTCT	GGTCTTCCGG	4320
	GGACACGCGG	CGGGTCGGGT	GTCCCCGGAC	CAATTAATCC	CCCCAACAAC	GATGTTGTTT	4380
	TCCCCGGAGG	TTCCCCCGTG	GCTCAATATT	GTTATGCCTA	TCCCCGGTTG	GACGATCCCG	4440
	GGCCCTTGGG	TTCCGCGGAC	GCCGGGCGGC	AAGACCTGCC	CCGGCGCGTC	GTCCGTCACG	4500
20	AGCCCTTGGG	CCGCTCGTTC	CTCACGGGGG	GGCTGGTTTT	GCTGGCGCCG	CCGGTACGCG	4560
	GATTGGCGC	ACCCAACGCA	ACGTATGCGG	CCCGTGTGAC	GTACTACCGG	CTCACCCGCG	4620
	CCTGCCGTCA	GCCCATCCTC	CTTCGGCAGT	ATGGAGGGTG	TCGCGGCGGC	GAGCCGCCGT	4680
	CCCCAAAGAC	GTGCGGGTCG	TACACGTACA	CGTACCAGGG	CGGCGGGCCT	CCGACCCGGT	4740
	ACGCTCTCGT	AAATGCTTCC	CTGCTGGTGC	CGATCTGGGA	CCGCGCCGCG	GAGACATTCG	4800
25	AGTACCAGAT	CGAACTCGGC	GGCGAGCTGC	ACGTGGGTCT	GTGTGTTGGTA	GAGGTGGGCG	4860
	GGGAGGGCCC	CGGCCCCACC	GCCCCCCCAC	AGGCGGCGCG	TGCGGAGGGC	GGCCCCGTGC	4920
	TCCCCCGGTT	CCCCGCGGGC	CGCCCGTGGC	GCTCGGTGCC	CCCGGTATGG	TATTCCGCCC	4980
	CCAACCCCGG	GTTTCGTGGC	CTGCGTTTCC	GGGAGCGCTG	TCTGCCCCCA	CAGACGCCCG	5040
	CCGCCCCCAG	CGACCTACCA	CGCGTCGCTT	TTGTCCCCCA	GAGCCTGCTG	GTGGGGATTA	5100
30	CGGGCCGCAC	GTTTATTTCG	ATGGCACGAC	CCACGGAAGA	CGTCGGGGTC	CTGCCACCCC	5160
	ATTGGGCCCC	CGGGGCCCTA	GATGACGGTC	CGTACGCCCC	CTTCCCACCC	CGCCCGCGGT	5220
	TTGCACGCGC	CCTGCGGACA	GACCCCGAGG	GGGTCGACCC	CGACGTTCCG	GCCCCCTTAA	5280
	CCGGGCGGCG	CCTCATGGCC	TTGACCGAGG	ACGCGTCCTC	CGATTGCGCT	ACGTCCGCTC	5340
	CGGAGAAGAC	GCCCCCTCCCT	GTGTGCGCCA	CCGCCATGGC	GCCCTCAGTC	GACCCAAGCG	5400
35	CGGAACCGAC	CGCCCCCGCA	ACCACTACTC	CCCCCGACGA	GATGGCCACA	CAAGCCGCAA	5460
	CGGTGCGCGT	TACGCCGGAG	GAAACGGCAG	TCGCCTCCCC	GCCCGCGACT	GCATCCGTGG	5520
	AGTCGTCGCC	ACTCCCCGCC	GCGGCGGCAA	CGCCCGGGGC	CGGGCACACG	AACACCAGCA	5580
	GCGCCCCCGC	AGCGAAAACG	CCCCCCACCA	CACGAGCCCC	CACGACCCCC	CCGCCCACGT	5640
	CTACCCACGC	GACCCCCCGC	CCCACGAGTC	CGGGGCCCCA	AACAACCCCT	CCCGGACCCG	5700
40	CAACCCCGGG	TCCGGTGGGC	GCCTCCGCCG	CACCCACGGC	CGATTCCCCC	CTCACC GCCT	5760
	CGCCCCCGCG	TACCGCGCCG	GGGCCCTCGG	CCGCCAACGT	TTCGGTGCCT	GCGACCACCG	5820
	CCACGCCCCG	AACCCGGGGC	ACCGCCCGTA	CCCCCCCCAA	GGACCCAAAG	ACGCACCCAC	5880
	ACGGACCCCG	GGACGCTCCC	CCCGGCTCGC	CAGCCCCCCC	ACCCCCCGAA	CATCGCGGCG	5940

	GACCCGAGGA	GTTTGAGGGC	GCCGGGGACG	GCGAACCCCC	CGATGACGAC	GACAGCGCCA	6000
	CCGCGCTCGC	CTTCCGAAC	CCGAACCCCA	ACAAACCACC	CCCCGCGCGC	CCCGGGCCCA	6060
	TCCGCCCCAC	GCTCCCGCCA	GGAATCTCTG	GGCCGCTCGC	CCCCAACACG	CCTCGCCCCC	6120
	CCGCCCAAGC	TCCCGCTAAG	GACATGCCCT	CGGGCCCCAC	ACCCCAACAC	ATCCCCCTGT	6180
5	TCTGGTTCCT	AACGGCCTCC	CCTGCTCTAG	ATATCCTCTT	TATCATCAGC	ACCACCATCC	6240
	ACACGGCGGC	GTTTCGTTGT	CTGGTCGCCT	TGGCAGCACA	ACTTTGGCGC	GGCCGGGCGG	6300
	GGCGCAGGCG	ATACGCGCAC	CCGAGCGTGC	GTTACGTATG	TCTGCCACCC	GAGCGGGATT	6360
	AGGGGGTGGG	GTGGGGGCGA	GAAACGATGA	AGGACGGGAA	AGGGAACAGC	GACCAAATGC	6420
	CACGATAAGA	ACAATAAACC	TGTGACGTCA	ATCGGATATG	TGAGTTTGGT	TGTGTTTTGT	6480
10	GGGACTGGGG	GCGGGGGGTG	GGAGGTATCA	GTGGGTGACA	GAGTCTTTTA	AAAGACGTGT	6540
	CCCGGGGCCC	TCGAGACGCG	CAACTTTTGG	CCACACAGAG	AAAGGCCCCC	AGACGAAGTC	6600
	ACCCGGGTCC	CCGAACAAAA	ACAAAAACCT	TGACCGCCGC	CGGGGGGCGT	GCCTGTTGTT	6660
	TTGGTCTCAA	TGGATCGGTA	TGCCGTTCGG	ACCTGGGGGA	TTGTGGGAAT	CCTCGGGTGT	6720
	GCTGCTGTTG	GGCCCGCACC	CACCGGCCCC	GCGTCCGATA	CAACAAACGC	GACCGCACGC	6780
15	CTCCCCACGC	ACCCCCACT	CATCCGTTCC	GGGGGCTTTG	CCGTCCCCCT	CATCGTGGGG	6840
	GGGCTGTGTC	TCATGATTCT	GGGGATGGCG	TGTCTACTCG	AGGTCCTGCG	TCGCTGGGT	6900
	CGCGAGTTGG	CGAGGTGCTG	CCCCACGCG	GGCCAATTTG	CCCCATGATT	TTTCGCCTTT	6960
	CTGGCCTTGC	CCCCACCCCA	TCGCCCCGAT	TGTGTGTCGG	GTGCCCCGGG	TACAGCAGCT	7020
	ATGGAGCGGT	CGGTAATATA	ACTTTGGTTG	TCGCCACACG	CCCCGTGCCG	GGCATGGGTT	7080
20	GTGCGGGAAA	GACGAAATAA	TCCGGCGATC	CCCAAGCGTA	CCAACTTGGG	GGGGGGGGGA	7140
	AAGAACTAA	AAACACATCA	AGCCCACAAC	CCATCCACAC	AGGGGGGTTA	TGGCGGACCC	7200
	ACCGCACCAC	CATACTCCGA	TTCGACCACA	TATGCAACCA	AATCACCCCC	AGAGGGGAGG	7260
	TTCCATTTTT	ACGAGGAGGA	GGAGTATAAT	AGAGTCTTTG	TGTTTAAAC	CCGGGGTCGG	7320
	TGTGGTGTTC	GGTCATAAGC	TGCATTGCGA	ACGACTAGTC	GCCGTTTTTC	GTGTGCATCG	7380
25	CGTATCACGG	CATGGGGCGT	TTGACCTCCG	GCGTCGGGAC	GGCGGCCCTG	CTAGTTGTCTG	7440
	CGGTGGGACT	CCGCGTCGTC	TGCGCCAAAT	ACGCCCTAGC	AGACCCCTCG	CTTAAGATGG	7500
	CCGATCCCAA	TCGATTTTCG	GGGAAGAACC	TTCCGGTTTT	GGACCAGCTG	ACCGACCCCC	7560
	CCGGGGTGAA	GCGTGTTTAC	CACATTTCAGC	CGAGCCTGGA	GGACCCGTTC	CAGCCCCCCA	7620
	GCATCCCGAT	CACTGTGTAC	TACGCAGTGC	TGGAACGTGC	CTGCCGCAGC	GTGCTCTAC	7680
30	ATGCCCCATC	GGAGGCCCCC	CAGATCGTGC	GCGGGGCTTC	GGACGAGGCC	CGAAAGCACA	7740
	CCTACAACCT	GACCATCGCC	TGGTATCGCA	TGGGAGACAA	TTGCGCTATC	CCCATCACGG	7800
	TTATGGAATA	CACCGAGTGC	CCCTACAACA	AGTCGTTGGG	GGTCTGCCCC	ATCCGAACGC	7860
	AGCCCCGCTG	GAGCTACTAT	GACAGCTTTA	GCGCCGTCAG	CGAGGATAAC	CTGGGATTCC	7920
	TGATGCACGC	CCCCGCCTTC	GAGACCGCGG	GTACGTACCT	GCGGCTAGTG	AAGATAAACG	7980
35	ACTGGACGGA	GATCACACAA	TTTATCCTGG	AGCACCGGGC	CCGCGCCTCC	TGCAAGTACG	8040
	CTCTCCCCCT	GCGCATCCCC	CCGGCAGCGT	GCCTCACCTC	GAAGGCCTAC	CAACAGGGCG	8100
	TGACGGTCGA	CAGCATCGGG	ATGCTCCCCC	GCTTTATCCC	CGAAAACCAG	CGCACCGTCG	8160
	CCCTATACAG	CTTAAAAATC	GCCGGGTGGC	ACGGCCCCAA	GCCCCCGTAC	ACCAGCACCC	8220
	TGCTGCCGCC	GGAGCTGTCC	GACACCACCA	ACGCCACGCA	ACCCGAACCT	GTCCCGGAAG	8280
40	ACCCCGAGGA	CTCGGCCCTC	TTAGAGGATC	CCGCCGGGAC	GGTGTCTTCG	CAGATCCCCC	8340
	CAAACCTGGCA	CATCCCGTCG	ATCCAGGACG	TCGCGCCGCA	CCACGCCCCC	GCCGCCCCCA	8400
	GCAACCCGGG	CCTGATCATC	GGCGCGCTGG	CCGGCAGTAC	CCTGGCGGTG	CTGGTCATCG	8460
	GCGGTATTGC	GTTTTGGGTA	CGCCGCCGCG	CTCAGATGGC	CCCCAAGCGC	CTACGTCTCC	8520

	CCCACATCCG	GGATGACGAC	GCGCCCCCT	CGCACCAGCC	ATTGTTTTAC	TAGAGGAGTA	8580
	TCCCCGCTCC	CGTGACCTC	TGGGCCCCGTG	TGGGAGGGTG	GCTGGGGTAT	TTGGGTGGGA	8640
	CTTGGACTCC	GCATAAAGGG	AGTCTCGAAG	GAGGGAAACT	AGGACAGTTC	ATAGGCCGGG	8700
	AGCGTGGGGC	GCGCACCGCT	GTCCCGACGA	TTAGCCACCG	CGCCACAGT	CACCTCGACC	8760
5	CGTCCGATCC	CGGTATGCCC	GGCCGCTCGC	TGCAGGGCCT	GGCGATCCTG	GGCCTGTGGG	8820
	TCTGCGCCAC	CGGCCTGGTC	GTCCGCGGCC	CCACGGTCAG	TCTGGTCTCA	GACTCACTCG	8880
	TGGATGCCCG	GGCCGTGGGG	CCCCAGGGCT	TCGTGGAAGA	GGACCTGCGT	GTTTTCGGGG	8940
	AGCTTCATTT	TGTGGGGGCC	CAGGTCCCC	ACACAACTA	CTACGACGGC	ATCATCGAGC	9000
	TGTTTCACTA	CCCCCTGGGG	AACCACTGCC	CCCGCGTTGT	ACACGTGGTC	ACACTGACCG	9060
10	CATGCCCCCG	CGCCCCGCC	GTGGCGTTCA	CCTTGTGTCTG	CTCGACGCAC	CACGCCACCA	9120
	GCCCCGCTTA	TCCGACCTTG	GAGCTGGGTC	TGGCGCGGCA	GCCGCTTCTG	CGGGTTCGAA	9180
	CGGCAACGCG	CGACTATGCC	GGTCTGTATG	TCCTGCGCGT	ATGGGTGCGC	AGCGCGACGA	9240
	ACGCCAGCCT	GTTTGTTTTG	GGGGTGGCGC	TCTCTGCCAA	CGGGACGTTT	GTGTATAACG	9300
	GCTCGGACTA	CGGCTCCTGC	GATCCGCGCG	AGCTTCCCTT	TTCGGCCCCG	CGCCTGGGAC	9360
15	CCTCGAGCGT	ATACACCCCC	GGAGCTCCC	GGCCACCCC	TCCACGGACA	ACGACATCCC	9420
	CGTCTCCCC	CCGAGACCCG	ACCCCCGCC	CCGGGACAC	AGGGACGCCC	GCGCCCCGCA	9480
	GCGGCGAGAG	AGCCCCGCC	AATTCCACGC	GATCGGCCAG	CGAATCGAGA	CACAGGCTAA	9540
	CCGTAGCCCA	GGTAATCCAG	ATCGCCATAC	CGGCGTCCAT	CATCGCCTTT	GTGTTTCTGG	9600
	GCAGCTGTAT	CTGCTTCATC	CATAGATGCC	AGCGCCGATA	CAGGCGCCCC	CGCGGCCAGA	9660
20	TTTACAACCC	CGGGGGCGTT	TCCTGCGCGG	TCAACGAGGC	GGCCATGGCC	CGCCTCGGAG	9720
	CCGAGCTGCG	ATCCCACCCA	AACACCCCC	CCAAACCCCG	ACGCCGTTCC	TCGTCTGCCA	9780
	CGACCATGCC	TTCCCTAACG	TCGATAGCTG	AGGAATCGGA	GCCAGGTCCA	GTCGTGCTGC	9840
	TGTCGCTCAG	TCCTCGGCCC	CGCAGTGGCC	CGACGGCCCC	CCAAGAGGTC	TAGGTCCAAG	9900
	CGGGCCGTTT	GGCAGGCCCG	CCCCACCGCC	CCCATCGTGG	TTATTTCCCC	CCCAATAAAC	9960
25	CGATGTTATT	TGCCTATATG	CGTGTGTTGG	ATCCCTTTGT	GATCGTTCGT	CATTCCCCGG	10020
	ATGGCATGGG	AGGCGGGTAA	TGGATGGGCG	GGGCCCCGGG	GGGGAGGAAA	AAGAATAAAG	10080
	GGGGTAGTGT	CGGAGAGGCC	CGCCGCGCAT	TTAAGGAGTC	GCCGCCCCGA	CTCTGTGTCT	10140
	TCGGGTGACT	TGGTGCGCCG	CCGTAGCTA	GTCTCCGATC	TGCCCCGACC	GACGGCTCCT	10200
	GCCACCCGAA	CATGGCTCGC	GGGGCCGGGT	TGTTGTTTTT	TGTTGGAGTT	TGGGTGCTAT	10260
30	CGTGCTTGGC	GGCAGCACCC	AGAACGTCCT	GGAAACGGGT	AACCTCGGGC	GAGGACGTGG	10320
	TGTTGCTTCC	GGCGCCCGCG	GGGCCGAGG	AACGCACCCG	GGCCACAAA	CTACTGTGGG	10380
	CCGCGGAACC	CCTGGATGCC	TGCGGTCCCC	TGCGCCCGTC	GTGGGTGGCG	CTGTGGCCCC	10440
	CCCGACGGGT	GCTCGAGACG	GTCGTGGATG	CGGCGTGCAT	GCGCGCCCCG	GAACCGCTCG	10500
	CCATAGCATA	CAGTCCCCCG	TTCCCCCGCG	GCGACGAGGG	ACTGTATTCTG	GAGTTGGCGT	10560
35	GGCGCGATCG	CGTAGCCGTG	GTCAACGAGA	GTCTGGTCAT	CTACGGGGCC	CTGGAGACGG	10620
	ACAGCGGTCT	GTACACCTTG	TCCGTGGTCG	GCCTAAGCGA	CGAGGCGCGC	CAAGTGGCGT	10680
	CGGTGGTTCT	GGTCGTGGAG	CCCGCCCTTG	TGCCGACCCC	GACCCCCGAC	GACTACGACG	10740
	AAGAAGACGA	CGCGGGCGTG	AGCGAACGCA	CGCCGGTCAG	CGTTCCCCCC	CCAACCCCCC	10800
	CCCGTGGTCC	CCCCGTGGCC	CCCCGACGC	ACCCTCGTGT	TATCCCCGAG	GTGTCCCACG	10860
40	TGCGCGGGGT	AACGGTCCAT	ATGGAGACCC	CGGAGGCCAT	TCTGTTTGCC	CCCGGGGAGA	10920
	CGTTTGGGAC	GAACGTCTCC	ATCCACGCCA	TTGCCACGCA	CGACGGTCCG	TACGCCATGG	10980
	ACGTCGTCTG	GATGCGGTTT	GACGTGCCGT	CCTCGTGCGC	CGAGATGCGG	ATCTACGAAG	11040
	CTTGCTCTGA	TCACCCGCAG	CTTCCAGAGT	GTCTATCTCC	GGCCGACGCG	CCGTGCGCCG	11100

	TAAGTTCTCTG	GGCGTACCGC	CTGGCGGTCC	GCAGCTACGC	CGGCTGTTCC	AGGACTACGC	11160
	CCCCGCCGCG	ATGTTTTGCC	GAGGCTCGCA	TGGAACCGGT	CCCGGGGTTG	GCGTGGCTGG	11220
	CCTCCACCGT	CAATCTGGAA	TTCCAGCACG	CCTCCCCCCA	GCACGCCGGC	CTCTACCTGT	11280
	GCGTGGTGTA	CGTGGACGAT	CATATCCACG	CCTGGGGCCA	CATGACCATC	AGCACCGCGG	11340
5	CGCAGTACCG	GAACGCGGTG	GTGGAACAGC	ACCTCCCCCA	GCGCCAGCCC	GAGCCCGTCG	11400
	AGCCACCCCG	CCCGCACGTG	AGAGCCCCC	CTCCCGCGCC	CTCCGCGCGC	GGCCCGCTGC	11460
	GCCTCGGGGC	GGTGCTGGGG	GCGGCCCTGT	TGCTGGCCCG	CCTCGGGCTG	TCCGCGTGGG	11520
	CGTGCAATGAC	CTGCTGGCGC	AGGCGCTCCT	GGCGGGCGGT	TAAAAGCCCG	GCCTCGGCGA	11580
	CGGGCCCCAC	TTACATTCGC	GTGGCGGACA	GCGAGCTGTA	CGCGGACTGG	AGTTCCGGACA	11640
10	GCGAGGGGGA	GCGCGACGGG	TCCCTGTGGC	AGGACCCTCC	GGAGAGACCC	GACTCTCCCT	11700
	CCACAAATGG	ATCCGGCTTT	GAGATCTTAT	CACCAACGGC	TCCGTCTGTA	TACCCCCATA	11760
	GCGAGGGGCG	TAAATCTCGC	CGCCCGCTCA	CCACCTTTGG	TTGGGGAAGC	CCGGGCGGTC	11820
	GTCACCTCCA	GGCCTCCTAT	TCGTCCGTCC	TCTGGTAAGG	CGTCTTCCGA	CGACGCGGAC	11880
	GTCGGCGATG	AACTGATTGC	CATCGCGGAC	GCACGCGGGG	ACCCGCCAGA	GACCTGCCCC	11940
15	CCCGGCGCGG	GCGGCGCCGC	GCCCGCGTGC	CGCAGACCAC	CTCGCGGCGG	CTCCCCCGCG	12000
	GCCTTTCCCG	TGGCCCTCCA	CGCCGTGGAC	GCCCCCTCCC	AATTTCGTAC	CTGGCTCGCC	12060
	GTGCGCTGGC	TGCGGGGGGC	GGTGGGTCTC	GGGGCCGTCC	TGTGCGGGAT	TGCGTTTTAC	12120
	GTGACGTCAA	TCGCCCAGAG	CGCATAAAGG	TCCGGCGGCC	AGCCCCGCCG	CAGCTCATAA	12180
	AAATCGTGAG	TCACGGCAAC	CGCACCTTCG	CCTCCGGCCC	TCCGCCAGCG	CCCTTCCGCG	12240
20	TCCGCGATGA	CCTCCCGGCC	CGCCGACCAA	GACTCGGTGC	GTTCCAGCGC	GTGCGTGCCG	12300
	CTTTACCCCG	CGGCCTCGCC	CGTCCCGGCA	GAAGCCTACT	ACTCGGAAAG	CGAAGACGAG	12360
	GCCGCCAACG	ACTTCCTCGT	GCGCATGGGC	CGCCAGCAGT	CGGTCTTAAG	GCGCCGACGG	12420
	CGGCGCACGC	GGTGCGTCGG	GCTGGTTATC	GCCTGTCTCG	TCGTGGCCCT	CCTATCTGGA	12480
	GGGTTCGGGG	CACTTTTGGT	GTGGCTGCTC	CGCTAAATGA	CGCCTCGATG	TATGGCGCCT	12540
25	TCTTCGCCCC	CACCCCTCGC	CGCGACCCAC	GTCCGTATGT	TAATTGCAAT	AAAGTGGTTG	12600
	ATTGTCATTA	CGGTCTACTA	GGTTGTCTTT	TTTTTTTGGG	GGGGGGGGGA	AGGAAATGCA	12660
	GAAAAGGGTA	AGAAATTCTC	GGAATTTCAC	CCCCCGGGGG	GGGCAAGTGC	AGTACCCACG	12720
	TTCTCTAGTG	TTTGGGAAAT	CTATTGAACT	CTCCCGGCTC	CTCCGTGTTA	GGGAAGTCTC	12780
	TTGGGGAAAT	CTATTGACCT	CTCGCCCCC	CCCCCAGGAG	GGGGCAGTGC	AGTACCCACG	12840
30	TTCTTCCGTG	CTGGGGAAAT	CTCTCTGCCG	GGTACGGGCT	CCAGACGAAG	GACCCATACA	12900
	TTTCCCCATC	CGCACCCAC	ATCTGGCGTT	CTAGAGTCAC	GACGCATTTG	CCCCCGTCCC	12960
	CGCAGCAACA	CACAAAGCGA	TTTCAATTTT	CACGATTTTA	TTATTAATTA	CACCAACCAC	13020
	CCTGTCCCCG	GGACGTGGTC	AGGACCGGGG	GTCCGCACCC	AAACGCACGA	AACAAATGCT	13080
	GGCAGTGTGC	CGAATATAAC	CCCGCGTAGG	AACACGTGCA	CGCGTGCGCC	AAACAGCACC	13140
35	AGAAGGCGCA	TGCCATCAGC	AGGTCTGTGA	TATGGCGATG	TGTTTGGACG	CAGGGCGCAG	13200
	CCGCGGCGAT	AAAATTTCATG	GCGGCGGTCC	GCCAGGGCCA	CAGCGGCGAG	GACTCCCTGT	13260
	TGGCCCGAAG	CCATTGGGTA	TGAACCAGCT	GCGCCTCCTG	TCCGACCCTG	GCTCCCGCCA	13320
	GCGGGGGCGG	TGGGTCGTGG	GTGTTGAGAG	CACACAGGCG	GGACACCTCG	ATCACCGTCC	13380
	GAAAAAAGGC	CCGGTGGTCC	GCGGGCAGCA	TCTGCAGGTG	CGCCAGGGCC	TGGGCGTTGA	13440
40	GAGGGTACAA	CTCGGAGCCG	GGGGACTCCG	GGGGCCGGTC	CGCGCGGTGC	CGCGAGTGGG	13500
	CACGCTTTGG	GGCCCGGGTG	TCGGACGCGG	GCGCGTTACG	GATCCCGACG	CGGGGCAGAA	13560
	CGTACGTGCG	TTGGCGCGGC	GATGAGGGGT	CCGGGCTGCC	GAGGGGGGCG	TAGGGGACCG	13620
	GGCTAGGCAA	GCCCGCGGGT	TGCGCGGGGT	TCCCGTGGGG	GTCTAGGCTC	CCTGGGCACC	13680

	CGTGGGGGTC	GTGGGGGTCG	CGGGTCCCTG	GGTATGCGCG	GGACCCTGGG	TTC'TCTGGGA	13740
	GATCGTGGA	CTCGCGGTT	CCTGGGCTCT	CGGGGAACCC	GGGGCTCCCT	GGGGACACGT	13800
	GGTGCCCTGG	GAAT'TCTTGA	TGGTCGGACG	GCTTCAGATG	GCTTCGGGAT	CGAGAGGGCC	13860
	GCACAGACTC	GTAGTAGACC	CGAATCTCCA	CGTTTCCCCG	CCGCCGGATC	ATGGTCGCCG	13920
5	CCCCGGTGCG	GGGGCCCGTC	GGTCGGAAGC	GAGTGCCCTT	CAAGCGTGTC	CGCTCCTCTG	13980
	GGCTGCATGC	CGTCGGATGG	GGTGCCTTTT	AAGGAAAGGT	CTCGGCTGCC	CGCCCCAAC	14040
	GGGGTTTGGG	GGTGGGCCGG	GGAAACCCCG	GATGCCATGG	GGGGGGTCAC	ACCTTAAGCG	14100
	CCGGCGCGCT	GGTTGGGTGG	GGGTAGAGGG	GAGTCCCCGG	TCGACGAGAT	CGTATCAAGG	14160
	GGCCAGCACG	CGATCCTGCC	GCTCGTTCGA	TCTAGCACAC	CCACGGGTCT	GCTGTGTGGG	14220
10	ATTTTCGACTC	GCGGGATCCG	ATCGCACGTC	CGGAGGACAC	AGCAGCGGGA	GCTCCGGGTC	14280
	GGTCACCGCA	GTTCTGGCCG	CCTCTCGGTC	CTCCCGTTCC	CTTTTATGGA	TCTCCGCGCA	14340
	GACATCGCCA	TACGTCCGGT	GTGTGCACCG	CGAAGAATCC	AAAAACATGT	CCGTCTGTTT	14400
	CAGGGCCCAA	GACATGGTGT	CCCGTCCACG	AAGGCGGCGC	CCGGCCTGCG	AGAAAGCGCG	14460
	GATGTTGGGA	TCGGGGCCCC	CCCGTCCCGT	CCCCCGTCC	CGTCCCCCGG	TCCCGTCCCC	14520
15	CCGTCCCGTC	CCCCCGTCCC	GTCCCCCGGT	CCCGTCCCCG	CGTCCCGTCC	CCCCGTCCCC	14580
	TCCCCCGTCC	CCGTCCCCCG	GTCCCCGTCC	CCCGTCCCGT	CCCCCGTCC	CGTCCCCCGG	14640
	TCCCGTCCCC	CCGTCCCCGT	CCCCCGTCCC	GTCCCCCGGT	CCCGTCCCCG	CGCCCCGGCG	14700
	CCCCCGGGTC	ACCGTACCTG	CGATAAGGCT	GCAAGTGGTG	GATGGGTCCT	CGCGGTACGT	14760
	ACAGGGTGGG	GGGGGGGGGG	AGGGAAAGGC	AGAACGAAAA	GGAACCGATG	CGCCCGCGTC	14820
20	TCTGTATCCG	ATCCGATCCG	GGTGCCTCGG	TGCCCCGCTC	GCCGCCGGCG	TCTCTGTCTC	14880
	GCTGTGGCCC	CCTTCGCGAT	GCCGCCGCTG	CCGTCCCGGT	CTCCGCCGCG	CAGCCGGTGT	14940
	GCCCCCTGGT	CGGCGGCGAC	CGGGACGCCG	GCCCTTTATG	TGCGCGAGGA	ACGGCCCGCC	15000
	CCCCGTCCGG	GCCCGCCTCG	GGGCGGGGCC	CGCGGGATGA	CGCGGGCCCC	GGGCAGGGCG	15060
	CCAGTGCTCG	CACTTTGCCC	TAATAATATA	TATACTATTA	GGACGAAGTG	CGAACGCTTC	15120
25	GCGTTCTCAC	TTCTTTTACC	CTGCGGCCCC	GCCCCCTTTG	GGGCGGAGCC	CGCGGGATGA	15180
	CGCGGGCCCC	GGGCAGGGCG	CCAGTGCTCG	CACTTTGCCC	TAATAATATA	TATACTATTA	15240
	GGACGAAGTG	CGAACGCTTC	GCGTTCTCAC	TTCTTTTACC	CTGCGGCCCC	GCCCCCTTTG	15300
	GGGCGGAGCC	GCCCGCGGAC	CAACGGGGCG	ACCTCGCCGG	CCCCAAAGGG	GCCGGCGGGG	15360
	GCCAACGGGA	GCGCGGGGCC	GGCATCTCAT	TACCACGAAC	CCGGAAGGGC	AGGGGAGCGA	15420
30	GCCCCCCCCG	GACGAGGGTC	TCATTAGCAT	CGCGGGCGGA	AGCGGAAGCC	GCCCCGCGCG	15480
	GGCGCTAATG	AGATGCCGCG	CGGGCGGAGC	CGGCGGCGGC	GCGACCAACG	GGCCGCCGCC	15540
	ACGGACGCGG	ACGCGCGGGC	GTCGGGGCGG	GGCCGCGCAT	AATGCGGTTC	CACCTGGGGG	15600
	CGGAACCCCG	GCGAGCCGGG	GCGCGGCGGC	GTCGATCGCT	CCTCCTCCGC	GTCTCTCTCC	15660
	TTTCCCCCCG	CCCCGCGCGC	CCCGAGGACT	ATATCAGCCA	GGCGACGGGG	CGATCGTCCA	15720
35	CACGGAGCGC	GGCTACCGAC	GCGGCCGCCA	GGATCTACCC	GATCGGCGCG	GAGAGGCGAA	15780
	AAGACACAGG	CACACGCACG	CACCGCACGG	GGGGGAGAGA	GAGACCGCCA	ACCCCCCCCC	15840
	CCCCCCACTG	CCGCCCTGA	AGAAGAAGAA	GACCCCCCGC	ACACCCCGGT	CGGAGGCGAT	15900
	GTCGGCGGAG	CAGCGGAAGA	AGAAGAAGAC	GACGACGACG	ACGCAGGGCC	GCGGGGCCGA	15960
	GGTCGCGATG	GCGGACGAGG	ACGGGGGACG	TCTCCGGGCC	GCGGCGGAGA	CGACCGGCGG	16020
40	CCCCGGATCT	CCGGATCCAG	CCGACGGACC	GCCGCCACC	CCGAACCCCG	ACCGTCGCCC	16080
	CGCCGCGCGG	CCCGGGTTCG	GGTGGCACGG	TGGGCCGGAG	GAGAACGAAG	ACGAGGACGA	16140
	CGACGCGGCC	GCCGATGCCG	ATGCCGACGA	GGCGGCCCCG	GCGTCCGGGG	AGGCCGTCTGA	16200
	CGAGCCTGCC	GCGGACGGCG	TCGTCTCGCC	GCGGCAGCTG	GCCCTGCTGG	CCTCGATGGT	16260

```

GGACGAGGCC GTTCGCACGA TCCCGTCGCC CCCCCCGGAG CGCGACGGCG CGGAAGAAGA 16320
AGCAGCCCGC TCGCCTTCTC CGCCGCGGAC CCCCTCCATG TCGCCGATT ATGGCGAGGA 16380
GAACGACGAC GACGACGATG ACGAGGACCG CGACGCGGGC CGCTGGGTCC GCGGACCGGA 16440
GAACGACGTC CGCGGTCCGC GGGGCGTACC CGGACCCCAT GGCCAGCCTG TCGCCGCGAC 16500
5 CCCCCGGGCC CCGCCGACAC CACCACCACC ACCACCGCCG CCGCCGCGG CGCGCCCCC 16560
GCCGGCGCTC GACCGCCTCT GACTCATCAA AATCCGGATC CTCGTCGTCG GCGTCCTCCG 16620
CCTCCTCCTC CGCCTCCTCC TCCTCGTCTG CATCCGCCCTC CTCGTCTGAC GACGACGACG 16680
ACGACGCCGC CCGCGCCCCC GCCAGCGCCG CAGACCACGC CGCGGGCGGG ACCCTCGGCG 16740
CGGACGACGA GGAGGCGGGG GTGCCC CGGA GGGCCCCGGG GGCGGCGCCC CGGCCGAGCC 16800
10 CGCCCAGGGC CG 16813

```

## (2) INFORMATION FOR SEQ ID NO:205:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 414 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

```

Met Ala Asp Ile Pro Pro Asp Pro Pro Ala Leu Asn Thr Thr Pro Ala
25 1 5 10 15
Asn His Ala Pro Pro Ser Pro Pro Pro Gly Ser Arg Lys Arg Arg Arg
20 25 30
Pro Val Leu Pro Ser Ser Ser Glu Ser Glu Gly Lys Pro Asp Thr Glu
35 40 45
30 Ser Glu Ser Ser Ser Thr Glu Ser Ser Glu Asp Glu Ala Gly Asp Leu
50 55 60
Arg Gly Gly Arg Arg Arg Ser Pro Arg Glu Leu Gly Gly Arg Tyr Phe
65 70 75 80
Leu Asp Leu Ser Ala Glu Ser Thr Thr Gly Thr Glu Ser Glu Gly Thr
35 85 90 95
Gly Pro Ser Asp Asp Asp Asp Asp Ala Ser Asp Gly Trp Leu Val
100 105 110
Asp Thr Pro Pro Arg Lys Ser Lys Arg Pro Arg Ile Asn Leu Arg Leu
115 120 125
40 Thr Ser Ser Pro Asp Arg Arg Ala Gly Val Val Phe Pro Glu Val Trp
130 135 140
Arg Asn Asp Arg Pro Ile Arg Ala Ala Gln Pro Gln Ala Pro Ala Gln
145 150 155 160

```

	Ser	Ser	Gly	Asp	Arg	Ala	Ala	Ala	Pro	Arg	Arg	Ser	Ala	Arg	Gln	Ala
						165				170						175
	Gln	Met	Arg	Ser	Gly	Ala	Ala	Trp	Thr	Leu	Asp	Leu	His	Tyr	Ile	Arg
				180					185					190		
5	Gln	Cys	Val	Asn	Gln	Leu	Phe	Arg	Ile	Leu	Arg	Ala	Ala	Pro	Asn	Pro
				195				200					205			
	Pro	Gly	Ser	Ala	Asn	Arg	Leu	Arg	His	Leu	Val	Arg	Asp	Cys	Tyr	Leu
				210			215					220				
	Met	Gly	Tyr	Cys	Arg	Thr	Arg	Leu	Gly	Pro	Arg	Thr	Trp	Gly	Arg	Leu
10				225			230				235					240
	Leu	Gln	Ile	Ser	Gly	Gly	Thr	Trp	Asp	Val	Arg	Leu	Arg	Asn	Ala	Ile
					245					250					255	
	Arg	Glu	Val	Glu	Ala	Arg	Phe	Glu	Pro	Ala	Ala	Glu	Pro	Val	Cys	Glu
				260				265						270		
15	Leu	Pro	Cys	Leu	Asn	Ala	Arg	Arg	Tyr	Gly	Pro	Glu	Cys	Asp	Val	Gly
				275				280					285			
	Asn	Leu	Glu	Thr	Asn	Gly	Gly	Ser	Thr	Ser	Asp	Asp	Glu	Ile	Ser	Asp
				290			295					300				
	Ala	Thr	Asp	Ser	Asp	Asp	Thr	Leu	Ala	Ser	His	Ser	Asp	Thr	Glu	Gly
20				305			310				315					320
	Gly	Pro	Ser	Pro	Ala	Gly	Arg	Glu	Asn	Pro	Glu	Ser	Ala	Ser	Gly	Gly
					325					330					335	
	Ala	Ile	Ala	Ala	Arg	Leu	Glu	Cys	Glu	Phe	Gly	Thr	Phe	Asp	Trp	Thr
				340					345				350			
25	Ser	Glu	Glu	Gly	Ser	Gln	Pro	Trp	Leu	Ser	Ala	Val	Val	Ala	Asp	Thr
				355				360					365			
	Ser	Ser	Ala	Glu	Arg	Ser	Gly	Leu	Pro	Ala	Pro	Gly	Ala	Cys	Arg	Ala
				370			375					380				
	Thr	Glu	Ala	Pro	Glu	Arg	Glu	Asp	Gly	Cys	Arg	Lys	Met	Arg	Phe	Pro
30				385			390				395					400
	Ala	Ala	Cys	Pro	Tyr	Pro	Cys	Gly	His	Thr	Phe	Leu	Arg	Pro		
					405					410						

(2) INFORMATION FOR SEQ ID NO:206:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 414 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

40

## (ii) MOLECULE TYPE: peptide



## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

	Met	Ala	Asp	Ile	Pro	Pro	Asp	Pro	Pro	Ala	Leu	Asn	Thr	Thr	Pro	Ala	
	1				5					10					15		
5	Asn	His	Ala	Pro	Pro	Ser	Pro	Pro	Pro	Gly	Ser	Arg	Lys	Arg	Arg	Arg	
				20					25					30			
	Pro	Val	Leu	Pro	Ser	Ser	Ser	Ser	Glu	Ser	Glu	Gly	Lys	Pro	Asp	Thr	Glu
			35						40					45			
	Ser	Glu	Ser	Ser	Ser	Thr	Glu	Ser	Ser	Glu	Asp	Glu	Ala	Gly	Asp	Leu	
10		50					55					60					
	Arg	Gly	Gly	Arg	Arg	Arg	Ser	Pro	Arg	Glu	Leu	Gly	Gly	Arg	Tyr	Phe	
	65					70					75					80	
	Leu	Asp	Leu	Ser	Ala	Glu	Ser	Thr	Thr	Gly	Thr	Glu	Ser	Glu	Gly	Thr	
					85					90					95		
15	Gly	Pro	Ser	Asp	Asp	Asp	Asp	Asp	Asp	Ala	Ser	Asp	Gly	Trp	Leu	Val	
				100					105					110			
	Asp	Thr	Pro	Pro	Arg	Lys	Ser	Lys	Arg	Pro	Arg	Ile	Asn	Leu	Arg	Leu	
				115					120					125			
	Thr	Ser	Ser	Pro	Asp	Arg	Arg	Ala	Gly	Val	Val	Phe	Pro	Glu	Val	Trp	
20		130						135					140				
	Arg	Asn	Asp	Arg	Pro	Ile	Arg	Ala	Ala	Gln	Pro	Gln	Ala	Pro	Ala	Gln	
	145					150					155					160	
	Ser	Ser	Gly	Asp	Arg	Ala	Ala	Ala	Pro	Arg	Arg	Ser	Ala	Arg	Gln	Ala	
					165					170					175		
25	Gln	Met	Arg	Ser	Gly	Ala	Ala	Trp	Thr	Leu	Asp	Leu	His	Tyr	Ile	Arg	
				180					185					190			
	Gln	Cys	Val	Asn	Gln	Leu	Phe	Arg	Ile	Leu	Arg	Ala	Ala	Pro	Asn	Pro	
				195					200					205			
	Pro	Gly	Ser	Ala	Asn	Arg	Leu	Arg	His	Leu	Val	Arg	Asp	Cys	Tyr	Leu	
30		210					215						220				
	Met	Gly	Tyr	Cys	Arg	Thr	Arg	Leu	Gly	Pro	Arg	Thr	Trp	Gly	Arg	Leu	
	225					230					235				240		
	Leu	Gln	Ile	Ser	Gly	Gly	Thr	Trp	Asp	Val	Arg	Leu	Arg	Asn	Ala	Ile	
					245					250					255		
35	Arg	Glu	Val	Glu	Ala	Arg	Phe	Glu	Pro	Ala	Ala	Glu	Pro	Val	Cys	Glu	
				260					265					270			
	Leu	Pro	Cys	Leu	Asn	Ala	Arg	Arg	Tyr	Gly	Pro	Glu	Cys	Asp	Val	Gly	
				275					280					285			
	Asn	Leu	Glu	Thr	Asn	Gly	Gly	Ser	Thr	Ser	Asp	Asp	Glu	Ile	Ser	Asp	
40		290							295					300			
	Ala	Thr	Asp	Ser	Asp	Asp	Thr	Leu	Ala	Ser	His	Ser	Asp	Thr	Glu	Gly	
	305					310					315				320		
	Gly	Pro	Ser	Pro	Ala	Gly	Arg	Glu	Asn	Pro	Glu	Ser	Ala	Ser	Gly	Gly	

325 330 335  
 Ala Ile Ala Ala Arg Leu Glu Cys Glu Phe Gly Thr Phe Asp Trp Thr  
 340 345 350  
 Ser Glu Glu Gly Ser Gln Pro Trp Leu Ser Ala Val Val Ala Asp Thr  
 5 355 360 365  
 Ser Ser Ala Glu Arg Ser Gly Leu Pro Ala Pro Gly Ala Cys Arg Ala  
 370 375 380  
 Thr Glu Ala Pro Glu Arg Glu Asp Gly Cys Arg Lys Met Arg Phe Pro  
 385 390 395 400  
 10 Ala Ala Cys Pro Tyr Pro Cys Gly His Thr Phe Leu Arg Pro  
 405 410

## (2) INFORMATION FOR SEQ ID NO:207:

## 15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 287 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

25 Met Gly Val Val Val Val Ser Val Val Thr Leu Leu Asp Gln Arg Asn  
 1 5 10 15  
 Ala Leu Pro Arg Thr Ser Ala Asp Asp Ala Leu Trp Ser Phe Leu Leu  
 20 25 30  
 Arg Gln Cys Arg Ile Leu Ala Ser Glu Pro Leu Gly Thr Pro Val Val  
 30 35 40 45  
 Val Arg Pro Ala Asn Leu Arg Arg Leu Ala Glu Pro Leu Met Asp Leu  
 50 55 60  
 Pro Lys Phe Trp Ile Val Arg Thr Arg Ser Cys Arg Cys Pro Pro Asn  
 65 70 75 80  
 35 Thr Thr Thr Gly Leu Phe Ala Glu Asp Asp Pro Leu Glu Ser Ile Glu  
 85 90 95  
 Ile Leu Asp Ala Pro Ala Cys Phe Arg Leu Leu His Gln Glu Arg Pro  
 100 105 110  
 Gly Pro His Arg Leu Tyr His Leu Trp Val Val Gly Ala Ala Asp Leu  
 115 120 125  
 40 Cys Val Pro Phe Leu Glu Tyr Ala Gln Lys Thr Arg Leu Gly Phe Arg  
 130 135 140  
 Phe Ile Ala Met Lys Thr Asn Asp Ala Trp Val Gly Glu Pro Trp Pro

145                      150                      155                      160  
 Leu Pro Asp Arg Phe Leu Pro Glu Arg Thr Val Ser Trp Thr Pro Phe  
                                  165                      170                      175  
 Pro Ala Ala Pro Asn His Pro Leu Glu Asn Leu Leu Ser Arg Tyr Glu  
 5                      180                      185                      190  
 Tyr Gln Tyr Gly Val Val Val Pro Gly Asp Arg Glu Arg Ser Cys Leu  
                                  195                      200                      205  
 Arg Trp Leu Arg Ser Leu Val Ala Pro His Asn Lys Pro Arg Pro Ala  
                                  210                      215                      220  
 10 Ser Ser Arg Pro His Pro Ala Thr His Pro Thr Gln Arg Pro Cys Phe  
                                  225                      230                      235                      240  
 Thr Cys Met Gly Arg Pro Glu Ile Pro Asp Glu Pro Ser Trp Gln Thr  
                                  245                      250                      255  
 Gly Asp Asp Asp Pro Gln Asn Pro Gly Pro Pro Leu Ala Val Gly Asp  
 15                      260                      265                      270  
 Glu Trp Pro Pro Ser Ser His Val Cys Tyr Pro Ile Thr Asn Leu  
                                  275                      280                      285

## (2) INFORMATION FOR SEQ ID NO:208:

20

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 479 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

25

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

30

Met Ala Cys Arg Lys Phe Cys Gly Val Tyr Arg Arg Pro Asp Lys Arg  
   1                      5                      10                      15  
 Gln Glu Ala Ser Val Pro Pro Glu Thr Asn Thr Ala Pro Ala Phe Pro  
                                  20                      25                      30  
 35 Ala Ser Thr Phe Tyr Thr Pro Ala Glu Asp Ala Tyr Leu Ala Pro Gly  
                                  35                      40                      45  
 Pro Pro Glu Thr Ile His Pro Ser Arg Pro Pro Ser Pro Gly Glu Ala  
                                  50                      55                      60  
 Ala Arg Leu Cys Gln Leu Gln Glu Ile Leu Ala Gln Met His Ser Asp  
 40    65                      70                      75                      80  
 Glu Asp Tyr Pro Ile Val Asp Ala Ala Gly Ala Glu Glu Glu Asp Glu  
                                  85                      90                      95  
 Ala Asp Asp Asp Ala Pro Asp Asp Val Ala Tyr Pro Glu Asp Tyr Ala

		100		105		110	
		Glu Gly Arg Phe Leu Ser Met Val Ser Ala Ala Pro Leu Pro Gly Ala					
		115		120		125	
		Ser Gly His Pro Pro Val Pro Gly Arg Ala Ala Pro Pro Asp Val Arg					
5		130		135		140	
		Thr Cys Asp Ser Gly Lys Val Gly Ala Thr Gly Phe Thr Pro Glu Glu					
		145		150		155	160
		Leu Asp Thr Met Asp Arg Glu Ala Leu Arg Ala Ile Ser Arg Gly Cys					
		165		170		175	
10		Lys Pro Pro Ser Thr Leu Ala Lys Leu Val Thr Gly Leu Gly Phe Ala					
		180		185		190	
		Ile His Gly Ala Leu Ile Pro Gly Ser Glu Gly Cys Val Phe Asp Ser					
		195		200		205	
		Ser His Pro Asn Tyr Pro His Arg Val Ile Val Lys Ala Gly Trp Tyr					
15		210		215		220	
		Ala Ser Thr Asn His Glu Ala Arg Leu Leu Arg Arg Leu Asn His Pro					
		225		230		235	240
		Ala Ile Leu Pro Leu Leu Asp Leu His Val Val Ser Gly Val Thr Cys					
		245		250		255	
20		Leu Val Leu Pro Lys Tyr His Cys Asp Leu Tyr Thr Tyr Leu Ser Lys					
		260		265		270	
		Arg Pro Ser Pro Leu Gly His Leu Gln Ile Thr Ala Val Ser Arg Gln					
		275		280		285	
		Leu Leu Ser Ala Ile Asp Tyr Val His Cys Glu Gly Ile Ile His Arg					
25		290		295		300	
		Asp Ile Lys Thr Glu Asn Ile Leu Ile Asn Thr Pro Glu Asn Ile Cys					
		305		310		315	320
		Leu Gly Asp Phe Gly Ala Ala Cys Phe Val Arg Gly Cys Arg Ser Ser					
		325		330		335	
30		Pro Phe His Tyr Gly Ile Ala Gly Thr Ile Asp Thr Asn Ala Pro Glu					
		340		345		350	
		Val Leu Ala Gly Asp Pro Tyr Thr Gln Val Ile Asp Ile Trp Ser Ala					
		355		360		365	
		Gly Leu Val Ile Phe Glu Thr Ala Val His Thr Ala Ser Leu Phe Ser					
35		370		375		380	
		Ala Pro Arg Asp Pro Glu Arg Arg Pro Cys Asp Asn Gln Ile Ala Arg					
		385		390		395	400
		Ile Ile Arg Gln Ala Gln Val His Val Asp Glu Phe Pro Thr His Ala					
		405		410		415	
40		Glu Ser Arg Leu Thr Ala His Tyr Arg Ser Arg Ala Ala Gly Asn Asn					
		420		425		430	
		Arg Pro Ala Trp Trp Ala Trp Thr Arg Tyr Tyr Lys Ile His Thr Asp					
		435		440		445	

Val Glu Tyr Leu Ile Cys Lys Ala Leu Thr Phe Asp Ala Ala Leu Arg  
 450 455 460  
 Pro Ser Ala Ala Glu Leu Leu-Arg Leu Pro Leu Phe His Pro Lys  
 465 470 475

5

## (2) INFORMATION FOR SEQ ID NO:209:

## (i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 37 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

Val Gly Gly Leu Cys Leu Met Ile Leu Gly Met Ala Cys Leu Leu Glu  
 1 5 10 15  
 20 Val Leu Arg Arg Leu Gly Arg Glu Leu Ala Arg Cys Cys Pro His Ala  
 20 25 30  
 Gly Gln Phe Ala Pro  
 35

25

## (2) INFORMATION FOR SEQ ID NO:210:

## (i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 385 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

Met Gly Arg Leu Thr Ser Gly Val Gly Thr Ala Ala Leu Leu Val Val  
 1 5 10 15  
 40 Ala Val Gly Leu Arg Val Val Cys Ala Lys Tyr Ala Asp Pro Ser Leu  
 20 25 30  
 Lys Met Ala Asp Pro Asn Arg Phe Arg Gly Lys Asn Leu Pro Val Leu  
 35 40 45  
 Asp Gln Leu Thr Asp Pro Pro Gly Val Lys Arg Val Tyr His Ile Gln

	50		55		60	
	Pro Ser Leu Glu Asp	Pro Phe Gln Pro	Pro Ser Ile Pro	Ile Thr Val		
	65	70	75	80		
	Tyr Tyr Ala Val Leu	Glu Arg Ala Cys Arg	Ser Val Leu Leu	His Ala		
5		85	90	95		
	Pro Ser Glu Ala Pro	Gln Ile Val Arg Gly	Ala Ser Asp Glu	Ala Arg		
		100	105	110		
	Lys His Thr Tyr Asn	Leu Thr Ile Ala Trp	Tyr Arg Met Gly	Asp Asn		
		115	120	125		
10	Cys Ala Ile Pro Ile	Thr Val Met Glu Tyr	Thr Glu Cys Pro	Tyr Asn		
		130	135	140		
	Lys Ser Leu Gly Val	Cys Pro Ile Arg Thr	Gln Pro Arg Trp	Ser Tyr		
		145	150	155	160	
	Tyr Asp Ser Phe Ser	Ala Val Ser Glu Asp	Asn Leu Gly Phe	Leu Met		
15		165	170	175		
	His Ala Pro Ala Phe	Glu Thr Ala Gly Thr	Tyr Leu Arg Leu	Val Lys		
		180	185	190		
	Ile Asn Asp Trp Thr	Glu Ile Thr Gln Phe	Ile His Arg Ala	Arg Ala		
		195	200	205		
20	Ser Cys Lys Tyr Ala	Leu Pro Leu Arg Ile	Pro Pro Ala Ala	Cys Leu		
		210	215	220		
	Thr Ser Lys Ala Tyr	Gln Gln Gly Val Thr	Val Asp Ser Ile	Gly Met		
		225	230	235	240	
	Leu Pro Arg Phe Ile	Pro Glu Asn Gln Arg	Thr Val Ala Lys	Leu Lys		
25		245	250	255		
	Ile Ala Gly Trp His	Gly Pro Lys Pro Pro	Tyr Thr Ser Thr	Leu Leu		
		260	265	270		
	Pro Pro Glu Leu Ser	Asp Thr Thr Asn Ala	Thr Gln Pro Glu	Leu Val		
		275	280	285		
30	Pro Glu Asp Pro Glu	Asp Ser Ala Leu Leu	Glu Asp Pro Ala	Gly Thr		
		290	295	300		
	Val Ser Ser Gln Ile	Pro Pro Asn Trp His	Ile Pro Ser Ile	Gln Asp		
		305	310	315	320	
	Val Ala Pro His His	Ala Pro Ala Ala Pro	Ser Asn Pro Gly	Leu Ile		
35		325	330	335		
	Ile Gly Ala Gly Ser	Thr Leu Ala Val Leu	Val Ile Gly Gly	Ile Ala		
		340	345	350		
	Phe Trp Val Arg Arg	Arg Ala Gln Met Ala	Pro Lys Arg Leu	Arg Leu		
		355	360	365		
40	Pro His Ile Arg Asp	Asp Asp Ala Pro Pro	Ser His Gln Pro	Leu Phe		
		370	375	380		
	Tyr					
	385					

## (2) INFORMATION FOR SEQ ID NO:211:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 368 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 10 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

15	Met	Pro	Gly	Arg	Ser	Leu	Gln	Gly	Leu	Ala	Ile	Leu	Gly	Leu	Trp	Val
	1				5				10					15		
	Cys	Ala	Thr	Gly	Leu	Val	Val	Arg	Gly	Pro	Thr	Val	Ser	Leu	Val	Ser
				20					25					30		
	Asp	Ser	Leu	Val	Asp	Ala	Gly	Ala	Val	Gly	Pro	Gln	Gly	Phe	Val	Glu
				35					40					45		
20	Glu	Asp	Leu	Arg	Val	Phe	Gly	Glu	Leu	His	Phe	Val	Gly	Ala	Gln	Val
		50						55					60			
	Pro	His	Thr	Asn	Tyr	Tyr	Asp	Gly	Ile	Ile	Glu	Leu	Phe	His	Tyr	Pro
	65				70						75				80	
	Leu	Gly	Asn	His	Cys	Pro	Arg	Val	Val	His	Val	Val	Thr	Leu	Thr	Ala
25				85					90					95		
	Cys	Pro	Arg	Arg	Pro	Ala	Val	Ala	Phe	Thr	Leu	Cys	Arg	Ser	Thr	His
				100					105					110		
	His	Ala	His	Ser	Pro	Ala	Tyr	Pro	Thr	Leu	Glu	Leu	Gly	Leu	Ala	Arg
			115						120					125		
30	Gln	Pro	Leu	Leu	Arg	Val	Arg	Thr	Ala	Thr	Arg	Asp	Tyr	Ala	Gly	Val
			130						135					140		
	Leu	Arg	Val	Trp	Val	Gly	Ser	Ala	Thr	Asn	Ala	Ser	Leu	Phe	Val	Leu
	145					150					155			160		
	Gly	Val	Ser	Ala	Asn	Gly	Thr	Phe	Val	Tyr	Asn	Gly	Ser	Asp	Tyr	Gly
35				165					170					175		
	Ser	Cys	Asp	Pro	Ala	Gln	Leu	Pro	Phe	Ser	Ala	Pro	Arg	Leu	Gly	Pro
				180					185					190		
	Ser	Ser	Val	Tyr	Thr	Pro	Gly	Ala	Ser	Arg	Pro	Thr	Pro	Pro	Arg	Thr
			195						200					205		
40	Thr	Thr	Ser	Pro	Ser	Ser	Pro	Arg	Asp	Pro	Thr	Pro	Ala	Pro	Gly	Asp
		210						215					220			
	Thr	Gly	Thr	Pro	Ala	Pro	Ala	Ser	Gly	Glu	Arg	Ala	Pro	Pro	Asn	Ser
	225					230					235				240	

	Thr	Arg	Ser	Ala	Ser	Glu	Ser	Arg	His	Arg	Leu	Thr	Val	Ala	Gln	Val	
						245					250					255	
	Ile	Gln	Ile	Ala	Ile	Pro	Ala	Ser	Ile	Ile	Ala	Phe	Val	Phe	Leu	Gly	
						260					265					270	
5	Ser	Cys	Ile	Cys	Phe	Ile	His	Arg	Cys	Gln	Arg	Arg	Tyr	Arg	Arg	Pro	
						275					280					285	
	Arg	Gly	Gln	Ile	Tyr	Asn	Pro	Gly	Gly	Val	Ser	Cys	Ala	Val	Asn	Glu	
						290					295				300		
	Ala	Ala	Met	Ala	Arg	Leu	Gly	Ala	Glu	Leu	Arg	Ser	His	Pro	Asn	Thr	
10	305						310					315				320	
	Pro	Pro	Lys	Pro	Arg	Arg	Arg	Ser	Ser	Ser	Ser	Thr	Thr	Met	Pro	Ser	
							325					330				335	
	Leu	Thr	Ser	Ile	Ala	Glu	Glu	Ser	Glu	Pro	Gly	Pro	Val	Val	Leu	Leu	
						340					345				350		
15	Ser	Val	Ser	Pro	Arg	Pro	Arg	Ser	Gly	Pro	Thr	Ala	Pro	Gln	Glu	Val	
						355					360				365		

(2) INFORMATION FOR SEO ID NO:212:

20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 528 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## 25

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

30	Met	Arg	Ala	Gly	Leu	Val	Phe	Phe	Val	Gly	Val	Trp	Val	Val	Ser	Cys
	1				5					10					15	
	Leu	Ala	Ala	Ala	Pro	Arg	Thr	Ser	Trp	Lys	Arg	Val	Thr	Ser	Gly	Glu
				20					25					30		
	Asp	Val	Val	Leu	Leu	Pro	Ala	Pro	Ala	Gly	Pro	Glu	Glu	Arg	Thr	Arg
35			35					40						45		
	Ala	His	Lys	Leu	Leu	Trp	Ala	Ala	Glu	Pro	Leu	Asp	Ala	Cys	Gly	Pro
		50					55						60			
	Leu	Arg	Pro	Ser	Trp	Val	Trp	Pro	Pro	Arg	Arg	Val	Leu	Glu	Thr	Val
	65					70				75					80	
40	Val	Asp	Ala	Ala	Cys	Met	Arg	Ala	Pro	Glu	Pro	Leu	Ala	Ile	Ala	Tyr
					85					90					95	
	Ser	Pro	Pro	Phe	Pro	Ala	Gly	Asp	Glu	Gly	Ser	Glu	Leu	Ala	Trp	Arg
				100					105					110		



Asp Arg Val Ala Val Val Asn Glu Ser Leu Val Ile Tyr Gly Ala Leu  
 115 120 125  
 Glu Thr Asp Ser Gly Thr Leu Ser Val Val Gly Leu Ser Asp Glu Ala  
 130 135 140  
 5 Arg Gln Val Ala Ser Val Val Leu Val Val Glu Pro Ala Pro Val Pro  
 145 150 155 160  
 Thr Pro Thr Pro Asp Asp Tyr Asp Glu Glu Asp Asp Ala Gly Val Ser  
 165 170 175  
 Thr Pro Val Ser Val Pro Pro Pro Thr Pro Pro Arg Gly Pro Pro Val  
 10 180 185 190  
 Ala Pro Pro Thr His Pro Arg Val Ile Pro Glu Val Ser His Val Arg  
 195 200 205  
 Gly Val Thr Val His Met Pro Glu Ala Ile Leu Phe Ala Pro Gly Glu  
 210 215 220  
 15 Thr Phe Gly Thr Asn Val Ser Ile His Ala Ile Ala His Asp Asp Gly  
 225 230 235 240  
 Pro Tyr Ala Met Asp Val Val Trp Met Arg Phe Asp Val Pro Ser Ser  
 245 250 255  
 Cys Ala Glu Met Arg Ile Tyr Glu Ala Cys Leu Tyr His Pro Gln Leu  
 20 260 265 270  
 Pro Glu Cys Leu Ser Pro Ala Asp Ala Pro Cys Ala Val Ser Ser Trp  
 275 280 285  
 Ala Tyr Arg Leu Ala Val Arg Ser Tyr Ala Gly Cys Ser Arg Thr Thr  
 290 295 300  
 25 Pro Pro Pro Arg Cys Phe Ala Glu Ala Arg Met Glu Pro Val Pro Gly  
 305 310 315 320  
 Leu Ala Trp Leu Ala Ser Thr Val Asn Leu Glu Phe Gln His Asp Gln  
 325 330 335  
 His Ala Gly Leu Cys Val Val Tyr Val Asp Asp His Ile His Ala Trp  
 30 340 345 350  
 Gly His Met Thr Ile Ser Thr Ala Ala Gln Tyr Arg Asn Ala Val Val  
 355 360 365  
 Glu Gln His Leu Pro Gln Arg Gln Pro Glu Pro Val Glu Pro Trp His  
 370 375 380  
 35 Val Arg Ala Pro Pro Pro Ala Pro Ser Arg Pro Leu Arg Leu Gly Ala  
 385 390 395 400  
 Val Leu Gly Ala Ala Leu Leu Leu Ala Ala Leu Gly Leu Ser Ala Trp  
 405 410 415  
 Ala Cys Met Thr Cys Trp Arg Arg Arg Ser Trp Arg Ala Val Lys Ser  
 40 420 425 430  
 Arg Ala Ser Ala Thr Gly Pro Thr Tyr Ile Arg Val Ala Asp Ser Glu  
 435 440 445  
 Leu Tyr Ala Asp Trp Ser Ser Asp Ser Glu Gly Glu Arg Asp Gly Ser

450                      455                      460  
 Leu Trp Gln Asp Pro Pro Glu Arg Pro Asp Ser Pro Ser Thr Asn Gly  
 465                      470                      475                      480  
 Ser Gly Phe Glu Ile Leu Ser Pro Thr Ala Pro Ser Val Tyr Pro His  
 5                      485                      490                      495  
 Ser Glu Gly Arg Lys Ser Arg Arg Pro Leu Thr Thr Phe Gly Ser Gly  
                     500                      505                      510  
 Ser Pro Gly Arg Arg His Ser Gln Ala Ser Tyr Ser Ser Val Leu Trp  
                     515                      520                      525

10

## (2) INFORMATION FOR SEQ ID NO:213:

## (i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 41 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:

Val His Ala Val Asp Ala Pro Ser Gln Phe Val Thr Trp Leu Ala Val  
   1                      5                      10                      15  
 25 Arg Trp Leu Arg Gly Ala Val Gly Leu Gly Ala Val Leu Cys Gly Ile  
                     20                      25                      30  
 Ala Phe Tyr Val Thr Ser Ile Arg Ala  
                     35                      40

30

## (2) INFORMATION FOR SEQ ID NO:214:

## (i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 85 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:

Met Thr Ser Arg Pro Ala Asp Gln Asp Ser Val Arg Ser Ser Ala Ser  
   1                      5                      10                      15

Val Pro Leu Tyr Pro Ala Asp Val Pro Ala Glu Ala Tyr Tyr Ser Glu  
                             20                            25                            30  
 Ser Glu Asp Glu Ala Ala Asn Asp Phe Leu Val Arg Met Gly Arg Gln  
                             35                            40                            45  
 5 Gln Ser Val Leu Arg Arg Arg Arg Arg Arg Thr Arg Cys Val Gly Leu  
                             50                            55                            60  
 Val Ile Ala Cys Leu Val Val Leu Ser Gly Gly Phe Gly Ala Leu Leu  
                             65                            70                            75                            80  
 Val Trp Leu Leu Arg  
                             85  
 10

## (2) INFORMATION FOR SEQ ID NO:215:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 227 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

Met Ser Ala Glu Gln Arg Lys Lys Lys Lys Thr Thr Thr Thr Thr Gln  
 25 1                            5                            10                            15  
 Gly Arg Gly Ala Glu Val Ala Met Ala Asp Glu Asp Gly Gly Arg Leu  
                             20                            25                            30  
 Arg Ala Ala Ala Glu Thr Thr Gly Gly Pro Gly Ser Pro Asp Pro Ala  
                             35                            40                            45  
 30 Asp Gly Pro Pro Pro Thr Pro Asn Pro Asp Arg Arg Pro Ala Ala Arg  
                             50                            55                            60  
 Pro Gly Phe Gly Trp His Gly Gly Pro Glu Glu Asn Glu Asp Glu Asp  
                             65                            70                            75                            80  
 Asp Asp Ala Ala Ala Asp Ala Asp Ala Asp Glu Ala Ala Pro Ala Ser  
                             85                            90                            95  
 35 Gly Glu Ala Val Asp Glu Pro Ala Ala Asp Gly Val Val Ser Pro Arg  
                             100                            105                            110  
 Gln Leu Ala Leu Leu Ala Ser Met Val Asp Glu Ala Val Arg Thr Ile  
                             115                            120                            125  
 40 Pro Ser Pro Pro Pro Glu Arg Asp Gly Ala Glu Glu Glu Ala Ala Arg  
                             130                            135                            140  
 Ser Pro Ser Pro Pro Arg Thr Pro Ser Met Cys Ala Asp Tyr Gly Glu  
                             145                            150                            155                            160

[illegible]

(2) INFORMATION FOR SEO ID NO:216:

(i) SEQUENCE CHARACTERISTICS:

15                    (A) LENGTH: 227 amino acids  
                      (B) TYPE: amino acid  
                      (C) STRANDEDNESS: single  
                      (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

25	Met	Ser	Ala	Glu	Gln	Arg	Lys	Lys	Lys	Lys	Thr	Thr	Thr	Thr	Gln
	1				5					10				15	
	Gly	Arg	Gly	Ala	Glu	Val	Ala	Met	Ala	Asp	Glu	Asp	Gly	Gly	Leu
				20					25					30	
	Arg	Ala	Ala	Ala	Glu	Thr	Thr	Gly	Gly	Pro	Gly	Ser	Pro	Asp	Ala
				35				40						45	
30	Asp	Gly	Pro	Pro	Pro	Thr	Pro	Asn	Pro	Asp	Arg	Arg	Pro	Ala	Arg
	50						55				60				
	Pro	Gly	Phe	Gly	Trp	His	Gly	Gly	Pro	Glu	Glu	Asn	Glu	Asp	Asp
	65					70				75				80	
	Asp	Asp	Ala	Ala	Ala	Asp	Ala	Asp	Ala	Asp	Glu	Ala	Ala	Pro	Ser
35					85					90				95	
	Gly	Glu	Ala	Val	Asp	Glu	Pro	Ala	Ala	Asp	Gly	Val	Val	Ser	Arg
				100					105					110	
	Gln	Leu	Ala	Leu	Leu	Ala	Ser	Met	Val	Asp	Glu	Ala	Val	Arg	Ile
				115				120					125		
40	Pro	Ser	Pro	Pro	Pro	Glu	Arg	Asp	Gly	Ala	Glu	Glu	Glu	Ala	Arg
	130						135				140				
	Ser	Pro	Ser	Pro	Pro	Arg	Thr	Pro	Ser	Met	Cys	Ala	Asp	Tyr	Glu
	145					150					155				160

[illegible]

(2) INFORMATION FOR SEQ ID NO:217:

(i) SEQUENCE CHARACTERISTICS:

15                    (A) LENGTH: 10 base pairs  
                      (B) TYPE: nucleic acid  
                      (C) STRANDEDNESS: single  
                      (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:

	CGCCCGCGTT	TCTGCCGCC	GCGCGCTCCT	GTGTGGACCC	CGGGGTGGGC	GGCGGGGGGG	60
	GTGCCGTGGG	TGTGGCGGCG	GGGCGCGGGC	CGGGGCCGGG	GCTCGCTGGT	CTGCCGAAGT	120
25	AAAGAAAAGA	TCGCCACCGT	GTGTTCTGTCT	GTGTGTTCTG	CGCGGCGCCG	GGGCCCCCCT	180
	GCCGGGCGGG	GCGGTGGGGC	GGGGCCGGGG	TCGCGCGGGG	GAAGGAAGGA	AAGGCCCCCG	240
	AAGCGCCGGG	AGGGGGCGCC	GGCGCGACGC	GGGCGGCCGG	GCGGGGCGCG	GCGGCGGCCG	300
	GGCGGGGGCG	CGCGGCGGCC	GGGCGGGGGC	GCGCGGCGGC	CGGGCGGGGG	CGCGCTTTCC	360
	CCGCGTCGCC	CCTCGGGTTC	CCAAGACCTA	TCACGTGTGC	GCAGGGGAGG	GGAGGACGCG	420
30	GGGGAGGGGA	GGACGCGGGG	GAGGGGAGGA	CGCGGGGGAG	GGGAGGACGC	GGGGGAGGGG	480
	AGGACGCGGG	GGATATATAA	AGCGGTAGAA	AGCGCGGGAA	TGGGCATATT	GGACCCGCGT	540
	GATTCGGTTG	CTCGCGGTTG	TCTTGTTTTG	ACGTTTTTTA	TGCGGGAACA	AGGGGGCTTA	600
	CCGGTTACAC	TGTCCGCTCG	CTATGGGGTT	CGTCTGTCTG	TTTGGGCTTG	TCGTTATGGG	660
	AGCCTGGGGG	GCGTGGGGTG	GGTCACAGGC	AACCGAATAT	GTTCTTTCGTA	GTGTTATTGC	720
35	CAAAGAGGTG	GGGGACATAC	TAAGAGTGCC	TTGCATGCGG	ACCCCCGCGG	ACGATGTTTC	780
	TTGGCGCTAC	GAGGCCCCGT	CCGTTATTGA	CTATGCCCGC	ATAGACGGAA	TATTTCTTCG	840
	CTATCACTGC	CCGGGGTTGG	ACACGTTTTT	GTGGGATAGG	CACGCCCAGA	GGGCGTATCT	900
	GGTTAACCCC	TTTCTCTTTG	CGGCGGGATT	TTTGGAGGAC	TTGAGTCACT	CTGTGTTTCC	960
	GGCCGACACC	CAGGAAACAA	CGACGCGCCG	GGCCCTTTAT	AAAGAGATAC	GCGATGCGTT	1020
40	GGGCAGTCGA	AAACAGGCCG	TCAGCCACGC	ACCCGTCAGG	GCCGGTGTG	TAAACTTTGA	1080
	CTACTCACGC	ACTCGCCGCT	GCGTCGGGCG	ACGCGATTTA	CGGCCTGCCA	ACACCACGTC	1140
	AACGTGGGAA	CCGCCTGTGT	CGTCGGACGA	TGAAGCGAGC	TCGCAGTCGA	AGCCCCTCGC	1200
	CACCCAGCCG	CCCGTCCTCG	CCCTTTTCGAA	CGCCCCCCCCA	CGGCGGGTCT	CCCCGACGCG	1260

	AGGTCGGCGC	CGGCATACTC	GCCTCCGACG	CAACTAGCCA	CGTCTGCATC	GCAAGCCACC	1320
	CTGGGTCGGG	AGCAGGATAT	CCGACCCGTC	TAGCGGCCGG	GTCGGCTGTC	CAGCGTCGTC	1380
	GCCCTAGAGG	CTGTCCGCCG	GGCGTGATGT	TTTCCGCATC	TACGACCCCC	GAACAGCCCC	1440
	TGGGGCTGTC	GGGCGATGCG	ACGCCGCCCC	TGCCGACTTC	CGTGCCCCCTG	GACTGGGCCG	1500
5	CGTTTCGGCG	CGCGTTTCTG	ATCGACGACG	CCTGGCGGCC	CCTGTTGGAG	CCGGAGCTCG	1560
	CGAACCCCT	AACCGCGCGC	CTCCTCGCGG	AGTATGACCG	TCGGTGCCAG	ACCGAAGAGG	1620
	TGCTGCCGCC	GCGGGAGGAT	GTGTTCTCCT	GGACGCGGTA	TTGTACCCCC	GACGACGTGC	1680
	GCGTGTTTAT	CATCGGGCAG	GACCCGTACC	ACCATCCCCG	CCAGGCGCAC	GGCCTGGCGT	1740
	TTAGCGTGCG	TGCGGATGTG	CCGGTGCCCTC	CGAGTCTACG	GAACGTGCTG	GCGGCGGTAA	1800
10	AAAATTGTTA	CCCCGACGCG	CGCATGAGCG	GCCGCGGCTG	CCTGGAAAAG	TGGGCTCGCG	1860
	ACGGCGTGCT	GTTGTTGAAC	ACGACCCTGA	CCGTCAAGCG	CGGGGCGGCG	GCGTCCCACT	1920
	CCAAGCTTGG	ATGGGACCGC	TTTGTGGGCG	GGGTGGTCCG	ACGGCTGGCC	GCGCGCCGCC	1980
	CGGGCCTGGT	CTTTATGCTC	TGGGGCGCCC	ATGCCCAGAA	CGCGATCAGG	CCCAGCCCTC	2040
	GCCAACACTA	CGTCCTCAAG	TTTCTCACC	CGTCGCCCCCT	CTCCAAGGTC	CCGTTTGGGA	2100
15	CGTGCCAGCA	TTTCCTCGCC	GCGAATCGCT	ACCTCGAAAC	CCGGGACATT	ATGCCTATCG	2160
	ACTGGTCGGT	ATAAGATGCC	GACATCCGGG	GTCTTGATTT	ACGAGGGGGC	AATTAATAAA	2220
	GACTGTTGAT	GGTTAAATCT	CGGGTCTCAT	ACCGGTCCGT	GATGTCGGGC	GTGGGGGAAG	2280
	AGAGGGTCCC	CTCTGCGTTT	ACTATCCTTG	CCTCGTGGGG	CTGGACGTTT	GCACCCAGAG	2340
	ACCATGATCC	TGGCGCGTCG	CCGAATACGA	CGCCCATAGA	GTGATTGCG	GGGACCGCAC	2400
20	CGGACGCGCA	CGTGGGGCCT	CTCGACGGAG	AGCCGGACCG	GGATGCGATC	TCCCCTGTTA	2460
	CGTCGAGCGT	GGCCGGCGAC	CCGCCGGGGG	CGGACGGCCC	CTACGTCACC	TTTGATACTC	2520
	TGTTTATGGT	ATCTTCGATC	GACGAACTGG	GGCGCCGCCA	GCTCACGGAT	ACGATCCGTA	2580
	AGGACCTGCG	GCTGTCGCTG	GCCAAGTTCA	GCATCGCGTG	TACCAAGACC	TCGTGTTTTT	2640
	CGGGGACGGC	CGCGCGCCAG	CGCAAGCGCG	GAGCACCGCC	GCAACGCACA	TGCGTACCAC	2700
25	GCAGCAACAA	GAGCCTCCAG	ATGTTTCGTTT	TGTGCAAGCG	CGCCAACGCC	GCGCAGGTGC	2760
	GCGAGCAGCT	GCGGGCGGTT	ATTTCGGTCG	GCAAGCCGCG	CAAGTATTAC	ACGCGGTCCT	2820
	CGGATGGGCG	GCTCTGCCCC	GCCGTCCCCG	TGTTTGTACA	CGAGTTTGTT	TCGTCCGAAC	2880
	CCATGCGCCT	CCATCGAGAT	AACGTCATGC	TGTCTACGGA	ACCAGACTAA	GCACCCCCGC	2940
	CGTCCCCCTT	CTTTTCCCCC	TACCTTCC	CCCCTTACTG	ATGTGTTGTG	ACGTTTCAAT	3000
30	AAATAACACG	TAGCTTATTT	TGTTGGATGA	TGGATTGATT	GATTTTATTG	ACCGTTCGTT	3060
	CGCCCGGCGG	TGCCGTGCGC	GCGCGCAGAG	GGAATATGCA	AGCGGGCGGG	GTGGGGAGGA	3120
	AAGAAGGTTT	CAGGTTCCGG	GGGTGGGTC	TGCGTCGTCC	AGGGTGGGGC	TGATCTGAAT	3180
	TTCCCGCAGA	ACCTCGACCA	GTAGGTCTGT	TGTGTTTGCT	GGGAACTCGC	CCGCCGTTGG	3240
	GGATACGGGG	GCGGGGGGTG	TGGTTGGGGC	GACGTCCAGG	GGTGCGTTAT	CGCACCCCCG	3300
35	CGCCGCCCTCG	GGGGCCGTCC	CGTAGATCGT	TGCGGTGATG	TAGATGGTGT	CCGGGGTCCA	3360
	CACCACCGTC	AGGATGCCGG	CCGTGCGACT	CCGGACGCTT	TCGCCGTGCG	ATGAGCTGAC	3420
	CCAGGAGTCA	AAGGGGTACG	CGTACATATG	GGCGTCCAC	CAGCGCTCCA	GCCTCTGGGT	3480
	ACTAGCGCGT	CCTATAAAGC	GGTATGCGCA	AAATTCGGCA	CGACAGTCGA	TAATCACCAG	3540
	CAGCCCAGATG	GGGGTGTGTT	GTATCACCAC	GCCTCCGCGG	GGCAGGCGGT	CCTGGCGCGC	3600
40	TCGACCCCGC	GTCAGAACCG	GCGCGTCCC	TGACTCAAAC	ACGTGCACCA	CCTGTGCCGC	3660
	GTCCGGCAGC	GCGCTCGTTA	GCGACGCCCT	GGGGTGATGT	AGGCTGTACG	CGATGGTCGT	3720
	CTGGGGGTTT	CCCATGTCTC	GGGGGGGTGG	GGGTGAATGT	CACCCGGCCC	GGGTGCGGTG	3780
	GGAACGCGAG	GGAATGGAGG	GTTAATAGAC	AATGACCACA	TTCGGATCGC	GTAGAGCAGA	3840

	TAGTATGTGC	TCGCTAATGA	CGTCATCGCG	TTCGTGGCGC	TCCCGGAGCG	GGTTTAGATT	3900
	CATGTGCAGG	AACTCGGATG	AGGTGGTGCG	GGACATGGCT	ACGTACGCGC	TGTTTAGGCG	3960
	CAGGTTTCCG	GGCGTGAAGC	ATATGGCGAC	CTTGTCAGAG	CTGAGCCCCC	GGGAGCGCGT	4020
	GATGGTCATC	GCGAGTTTGG	AGCTGATGCC	GTAAGTCGGC	TTGATGGCCA	TGGCCAGCTC	4080
5	CGTGGAGTCG	ATCGACTCGA	CAAACCTACT	GATGTTGGTA	TTGACGACAG	ACATGAAGCC	4140
	GTGCTGGTCC	CGCAGGACGA	TGTAGGGCAG	GGGGGACTCC	TCCAAGAACT	CGGCCACGCC	4200
	GGCCGTCGCG	TGCCGCCGCC	GCAGCTCCTC	CGCGAACGCG	AACACCCGGG	TGTACGTGTA	4260
	CCCCATCAGC	GTGTAGTTGT	CCGTCTGCAG	GGCCACGGAC	ATCAGCCCCC	CGCGCGGCGA	4320
	GCCGGTCAGC	AGCTCGCAGC	CCCGGAAGAT	GACATTGTCC	ACGTAGGTGC	TGAAGGGGGC	4380
10	GCTCTCAAAC	ACCTCCCCGA	AGAGCTCCCC	TAGGATAAGG	TATCGCCCCA	GAAAGGCCCT	4440
	CTTCAGGAGC	CCAAACTGGG	CGTGGACGGC	CGCGGTGGTC	TCCGGCTCTT	CGAGGGCGTA	4500
	GTGGCAGTAG	AACACGTCCA	GCTGCTGTTT	TCCAGGCCCG	GCGAAGATAA	CGTCAAGGTC	4560
	GTCGTCGGGG	AAGTCGTCCG	GGCCCCCGTC	CCGCGGGCCC	AGGTGCTTAA	AATTGAACGC	4620
	ACGCTCCCCC	GGAGAGCGGT	CGCTGGTGTC	GGCGGCCCTG	GTTGCCGATG	CGCCGGCGGC	4680
15	GTCCCGGCGT	AGCGACAGGA	GTTCTGCCGT	CAGCTCCCCC	AGGCGGCCGT	AGGCCAGGGT	4740
	CCTCTGGGTC	GCGTCCAGGC	CGGGGCGCTG	GAGAAAGTTG	TAAAAGTGAA	TCAGCCCGCC	4800
	GAACATGAGC	CGCGACAGGA	ACCGGTAGGC	GAACTCCACC	GAGGTCTCCC	CCTGGGTCTT	4860
	CACGAAGCTG	TCGTGCGCGA	GCACAGCCTC	GAAAGTCCGA	AACGTCCCGT	CGAACCCAAA	4920
	CACCATCTTT	CGGAGGCGCG	CGGTCACCGC	GACCTGGCTG	TTGAGGACGT	ACGTGATGTC	4980
20	GTTCCGGGCC	ACGACTAGCT	GTTGCTTGCT	GTGCACCTCA	CAGCGCACGT	GCCCCGCGTC	5040
	CTGGTCTTGA	CTCTGGGAGT	AGTTGGTGAT	GCGACTGGCG	TTGGCCGTGA	TCCACTTTTC	5100
	CATGGTCAGC	GTGGGTTGCT	GCGTGAGCCG	TCGATACTCG	TCAAACCTCT	TGACCGACAC	5160
	AAACGTAAGC	ACGGGGAGGG	TAAACACAAC	AAACTCCCCC	TCGCGAGTCA	CCTTTAGGTA	5220
	GGCGTGGAGC	TTGGCCATGT	ACGCGCTGAC	CTCCTTGTGG	GACGAGAAAC	GCCGCGTCCA	5280
25	CCCCGGAAGG	TTGGCCGGGT	TGGTGATGTA	ACTTTCCGGG	ACGACAAAGC	GGTCCACAAA	5340
	CTGCATGTGC	TCCTCGGTGA	TGGGAAGGCC	GTAATCCAGC	ACCTTCATGA	GGTTCCCGAA	5400
	CTCGTGCTCC	ACACATCGCT	TGTTGTTAAT	GAAAATGGCC	CAGCTGTGCG	AGAGGCGCGT	5460
	GTAATCGCGT	AGGGTGCGGT	TGCAGATGAG	GTACGTGAGC	ACGTTTTCGC	TCTGCCGGAC	5520
	GGAGCATCGC	AGTTTTTGGT	GTTTGAAGGT	GGACTCCAGC	GAGGCCGTCT	GGGTGCGCGA	5580
30	CCCCACGCAC	ACCAGCACCG	GCCGCGGGCG	GCCCGCGTAC	TGGGGGGTGT	GGTACAGGGC	5640
	GTTAATCATC	CACCAGCAAT	ACACCACGGT	CGTGAGTAGG	TGCCGCCCCA	GGAGCCCGGC	5700
	CTCGTCGATG	ACGATAATGT	TGCTGCGGGT	GAAAGCCGGC	AGCGCCCCGT	GTGTGACCGA	5760
	GGCCAGGCGC	GTGAGGGCAC	CCTGGCCCAG	CCCCAAAGTC	TGCTCTAGGG	CGGTGAGGGC	5820
	GTGGAACTCG	TTTCGCGCGT	CTTCGCCCCC	GTGCGCCGCC	AGGGCCCGCT	TGGTGATGTC	5880
35	GAGGATCACC	TCCCAGTAGT	ACGTCAGGTC	TCGCGCGTGC	AGGTCCTCCA	GCGAGGCGGG	5940
	GCTGCTGGCC	AGGGTGACG	GGTGCTGCCC	CAGCTGGGCC	TGGACGTGAT	TCCCAGGAAA	6000
	CCCGAATCG	TGAAAGATGG	TGTTGATGGG	TCGACTCAGA	AACGCCCCCG	AGAGCTTAAC	6060
	GTACATGTTT	TGCGCCGCGA	TTCCGCTGGC	GCCCGTGACC	ACGAGTCCA	GGACCTCGTT	6120
	GAGGGTCTGC	ACGCACGTAC	TCTTTCCGGA	TCCGGCGTTG	CCGGTGATGA	GATACGCCGC	6180
40	GAACGGAAAC	TCCCGGAGCG	GCAGGCCGGT	CGGGACCTCC	AAGGCCGCGA	CGTCCCGGAA	6240
	CCACTGCAGG	CGCGGCACCT	GCGTGACGTC	GAGCTGCTGC	TGCGAGAGCT	CTCGGATGCG	6300
	TGCGATGATT	GGTTGACCC	CGTGACGGA	CGTAAATTTT	AAAAACGCCT	CGTCCCTGAA	6360
	CCGCACGGCG	GGTCTGGCCC	CGGGCTGCTG	TGGGGGCGGA	CCTGGTGCCC	GGACGTCCCG	6420

	CGAGCCCTCC	CCGCCGGACG	CCGCCATGGC	CGCACAGCGC	GCGCGGGCGC	CGGCGATGCG	6480
	GACGCGGGGC	GGCGACGCGG	CGCTATGCGC	CCCCGAGGAC	GGCTGGGTGA	AGGTTACACC	6540
	CACCCCGGG	ACGATGTTGT	TCCGCGAGAT	TCTCCTCGGG	CAGATGGGGT	ACACCGAGGG	6600
	TCAGGGGGTG	TACAACGTCG	TCCGGTCCAG	CGAGGCCGCC	ACCCGACAGC	TGCAGGCGGC	6660
5	GATCTTCCAC	GCGCTCCTCA	ACGCCACAAC	GTACCGGGAC	CTGGAGGAGG	ACTGGCGCCG	6720
	CCACGTGGTG	GCCCCGCGCC	TCCAGCCGCA	GCGGCTGGTT	CGCAGGTACC	GGAACGCCCC	6780
	GGAGGGCGAT	ATCGCCGGGG	TGGCCGAGCG	GGTGTTCGAC	ACGTGGCGAT	GCACGCTCAG	6840
	GACGACGCTG	CTGGACTTTG	CCCACGGGGT	GGTAGACTGC	TTTGCGCCGG	GCGGCCCAAG	6900
	CGGACCGACC	AGCTTCCCCA	AATATATCGA	CTGGCTGACG	TGTCTGGGGC	TGGTTCCCAT	6960
10	ATTGCGCAAG	ACGCGCGAGG	GGGAGGCGAC	GCAGCGCCTG	GGGGCGTTTC	TCAGGCAGCA	7020
	CACGCTGCCC	CGGCAGCTGG	CCACGGTCGC	CGGGGCCGCG	GAGCGCGCCG	GCCCCGGGCT	7080
	TCTGGAGCTG	GCCGTCGCGT	TCGACTCCAC	GCGCATGGCG	GAATACGACC	GTGTGCACAT	7140
	CTACTACAAC	CATCGCCGGG	GGGAGTGGCT	GGTGCAGCAG	CCGGTCAGCG	GGCAGCGCGG	7200
	CGAGTGCCCTG	GTGCTGTGCC	CCCCCTGTG	GACCGGCGAC	CGCCTGGTCT	TCGATTGCCC	7260
15	CGTTCAGCGG	CTGTGCCCCG	AGATCGTCGC	GTGCCACGCC	CTCCGGGAAC	ACGCGCACAT	7320
	CTGCCGTCTG	CGCAACACCG	CGTCCGTCAA	GGTGTGTTG	GGGCGCAAGA	GCGACAGCGA	7380
	GCGCGGGGTG	GCTGGCGCCG	CGCGGGTCGT	CAATAAGGCG	CTGGGGGAGG	ATGACGAGAC	7440
	GAAGGCCGGC	TCGGCCGCCT	CGCGTCTCGT	GCGGCTCATC	ATCAACATGA	AGGGCATGCG	7500
	CCACGTGGGC	GACATCAACG	ACACGGTACG	CGCCTACTTG	GACGAGGCGG	GGGGGCACCT	7560
20	GATCGACACC	CCCGCCGTCG	ACCACACCCT	CCCTGGGTTC	GGCAAGGGCG	GCACCGGCCG	7620
	CGGGTCGGCG	GCCCAGGACC	CGGGGGCGCG	ACCGCAGCAG	CTTCGCCAGG	CGTTTCAGAC	7680
	GGCCGTGGTC	AACAACATCA	ACGGCATGCT	GGAGGGCTAT	ATCAATAATC	TC'TTGGAAAC	7740
	CATAGAACGC	CTGCGAGAGA	CGAACCGGGG	TCTGGCGACC	CAGCTGCAGG	CGCGCGACCG	7800
	CGAGCTGCGG	CGCGCCAGG	CGGGGGCGCT	GGAGCGGGAG	CAGCGCGCGG	CGGACCGGGC	7860
25	GGCCGGGGGA	GGCGCGGGCC	GCCCGCGGGA	GGCGGATCTT	CTCCGGGCCG	ACTACGACAT	7920
	TATCGACGTC	AGCAAGTCCA	TGGACGACGA	CACGTACGTG	GCCAACAGTT	TCCAGCACCA	7980
	GTACATCCCC	GCGTACGGCC	AGGACCTCGA	GCGCCTGTG	CGCCTCTGGG	AGCACGAGCT	8040
	GGTGCCTGCT	TTCAAGATTG	TGCGCCACCG	CAACAACCAG	GGCCAGGAAA	CGTCGATCTC	8100
	GTACTCTAGC	GGGGCGATCG	CCTCCTTCGT	GGCCCCGTAT	TTGAGTACG	TGCTTCGCGC	8160
30	CCCCCGAGCG	GGCGCGCTCA	TCACCGGCTC	CGATGTATC	CTAGGGGAGG	AGGAGTTATG	8220
	GGAGGCGGTC	TTTAAGAAAA	CCCGCCTGCA	GACGTACCTG	ACAGACGTCG	CGGCCCTGTT	8280
	CGTGGCGGAC	GTACAGCACG	CGGCTCTGCC	CCGGCCCCCC	TCCCCAACCC	CCGCCGATTT	8340
	CCGGGCGAGC	GCGTCCCCGC	GGGGCGGGTC	CCGGTCCCGG	ACCCGAGCCC	GATCCCGGTC	8400
	GCCCCGGGAG	ACGCCGAGGG	GTGCGCCGGA	CCAGGGCTGG	GGCGTCGAAC	GCAGGGATGG	8460
35	CCGACCCAC	GCCCCCGCAT	GAGGGAACGG	CCGCCGCCAT	CCTCAAACAG	GCCATCGCCG	8520
	GGGACCGCAG	TCTGGTCGAG	GTGGCGGAGG	GGATCAGCAA	CCAGGCGCTG	CTGCGCATGG	8580
	CCTGCGAGGT	GCGCCAGGTC	AGCGATCGCC	AGCCGCGGTT	TACCGCGACC	AGCGTCCTGC	8640
	GCGTTGACGT	CACCCCCAGG	GGGCGGTTGC	GGTTGCTTCT	GGACGGGAGT	TCCGACGACG	8700
	CGTACGTGGC	GTCGGAGGAT	TACTTTAAGC	GCTGCGGGGA	CCAGCCGACG	TATCGCGGTT	8760
40	TTGCGGTCGT	CGTCCTCACG	GCCAACGAGG	ACCACGTGCA	CAGCCTGGCC	GTGCCCCCCC	8820
	TCGTTCTGCT	GCACCGGCTC	TCCTTGTTTC	GCCCCACGGA	CCTCCGGGAC	TTGAGCTCG	8880
	TCTGCCTGCT	GATGTACCTG	GAGAACTGTC	CCCGGAGCCA	CGCCACGCCC	TCGCTGTTTC	8940
	TCAAGGTGTC	GGCGTGTTG	GGGGTCGTGG	CCCCCACGCG	GTCTCCCTTC	GAGCGCGTCC	9000



	GCTGCCTTCT	CCTCCGCAGC	TGCCACTGGA	TCCTGAACAC	GCTAATGTGC	ATGGCGGGCG	9060
	TGAAGCCCTT	CGACGACGAG	CTAGTCCTGC	CCCCTGTA	CATGGCCAC	TACCTGCTGG	9120
	CCAACAATCC	GCCCCCGTC	CTCTCGGCC	TGTTTTGCGC	CACCCCGCAG	AGCTCTGCGT	9180
	TGCAGTTGCC	CGGGCCCGTC	CCCCGCACGG	ACTGTGTGGC	CTATAACCCG	GCCGGCGTCA	9240
5	TGGGAAGCTG	CTGGAAATCC	AAGGACCTGC	GTTCGGCTCT	GGTGTATTGG	TGGCTTTTCG	9300
	GGAGCCCCAA	ACGACGGACC	TCGTGCTTT	TCTATCGGTT	TTGCTAACTC	CGGAAAATAA	9360
	ACGTGTTTTT	TATGGAACGT	TCCCTACCTG	TCGTGTCATC	TCTCGGGGGA	TGGTGGTGGG	9420
	CCTGTGTGTG	TGTCTTGTGC	ACCGAAGGAG	GAAAGTGGGG	GGGTGGTGGT	GCTGGTGGTG	9480
	GAAAGACATG	ATAGAGGGAA	CAAAGAAATA	GAAGAAAACC	ACAACCGGCG	CGTGTACAGTA	9540
10	AATACGGACG	CGCGCACACG	CGGGGGTAAG	TTGGAGCACG	GGGCCCCAGT	TTATTGACCA	9600
	AATTCAAGGA	AACAGAAACC	GCATCTTTTC	CTCGAAAGGG	TACACAAAGC	TCCCGCCCTC	9660
	GCCCCACACG	CCTTCCAGAA	CCCCCGTAAA	CACCAAGTGA	ATCTCGCGCA	GGATCTCGCG	9720
	CAGGTGATGG	GCGCAGTCCA	CGGGGGGGAG	CACCAAGGGC	CGCGGGTACA	GATCCACGGG	9780
	GACGCCGACC	GACTCCCCGC	CCCCGGGACA	TACGCGCACG	ACGCGTCTCC	AGTATTGCTC	9840
15	CGCGTCCAGC	AGGGCGCCTC	CGCGGAAGGC	CGTTTGGGGC	AGGGGGTCGT	CGGCCTCGCC	9900
	CGGGGGGGTC	AGAACGCTCC	AGTACTCCGC	GTCCAGACGC	CTCCCGAAGG	CATCCAGGAC	9960
	AAAGCGGTCA	CAGGCGTCC	CCATGATGCC	CCGGGGCCGC	CACACGGCCT	CCTCCGGCGG	10020
	GCCGGCGGCC	GGCCGCCGGA	GGATTCTGTC	CAGCGCGTCG	CGCATAACCT	CGGCCGCCGC	10080
	GGCGTACGCG	GGCCCGCGGA	GAGGAAATCC	CTGCAGGAAG	TCGGTGTTCAT	CGCGGGAGTT	10140
20	CCAGAACCAC	GCCCCGGTCT	GGCTCCAGGT	GACGACGTGG	GTGTAGACGC	CCTCTAGCGC	10200
	CAGGGAGGGG	GCGAGGCGCG	GGCGTATGCC	GTTGGCCGAA	AGTACGGCGC	GCACGGACGC	10260
	CTCGAGGGCC	CGGCGGGCGT	CCTGGATCGC	GCCGTGCGCG	GCGTCCGCGT	CCCCGGGGTC	10320
	CACGTTGAAC	AGCCCCCAGA	ACGCAAGCCC	GGTGCCGCGC	CAGACCGCAA	ACTTCACCGA	10380
	GCTGGCCGTC	TGCTCGATCT	GCAGGCAGAC	GGCGGCCATG	ACCCCGCCGA	GCAGCTGCCG	10440
25	GAGCCCGGGG	CAGGCGTCGC	ACGCGTCCGG	CACCAGGCGC	TCCAGCACGG	CCCGGGCCCA	10500
	GGGCTCCGAG	GGGGCGGCCG	CCACCAGCGC	GTCCAGCCTT	TCCAGGCCCG	CCCGCCCCCG	10560
	GGCTTCCGGC	AGCCCGGCC	CCCCGAGGCC	CGCGAGGGCG	GCCAGGAGCT	GGGCCTGGAG	10620
	CCCCGAGAAA	CAAAACCGCG	CCGTCCAGAC	CGGCCCCGAC	GCCCGCGGGG	GGTCGAGTAG	10680
	TTGGATGGTG	GTGGCCGTGG	GGTGCCACCG	CGCCACCGCT	TCCCGAAAGG	CGGGCAGGAG	10740
30	GCGGCCGGCC	GCCTCCGAGG	CCACGGCCCG	CCATGCCCCG	GGGGGCAGGA	CGACCCTGGC	10800
	GCCCCACCGC	GGCCAGGCCC	CCAGGCACGC	GGCATGGGTG	GCCGCGGCGC	CCCGCACACG	10860
	GTCACGCGCC	GACTCGGCGG	CGGCGGCGGC	CGGCACGGTA	AACGTGGGCC	AGCCCCGAAA	10920
	TCCCAGCACG	GCAAAGTATT	GGACGGGCCC	TCCCCGGACC	TCAAACCCGG	GCCCCAGAAA	10980
	AGCGAAGACG	GGGGCCAGGG	CTCCGGGGGC	GGCGTGGACC	GTGGTATGCC	ACTGCCGGAA	11040
35	GAGGGCGACC	AGCGCCGGGG	CGGAGAACCC	GTGCGCGGCG	CTACGAAGT	AGTCGTAGCC	11100
	GCGCGGCAGC	AGCACCCGCG	CCGTGACCCG	CTGCGGGTGT	CCGCGGGGCC	GCAGGCCTAC	11160
	CTCGCACACC	TCGACCAGGT	CCGCGAAGGC	GCCCTCCTTG	CTGGTCGGCG	GAAACGCCAG	11220
	GGTGGTGTAT	TCGCGCGCGA	AACGCGCGGT	CCTCGTCGTG	ATGGTGACGG	CGAGCGAGGC	11280
	GGAGGACGCG	CACTGGGGGC	TGTCGCGAAT	GGCGGCCAGG	CGCGCCACG	CCAACGCGC	11340
40	GCCGGGGTGC	TCGGCGACGC	GCGCGGCCAG	GGCCAGCGGG	TCGACGTGCA	CCTTGGCCCTC	11400
	CACGTCCAGG	AGGGCGGCGC	GAGGAGCGGC	CGGCGGGCCC	CACGACGCCC	TTTCGACCCCT	11460
	CACGACCAGA	CCCGTCTGCG	GGTCCCAGCC	CAGGCGCAGC	GGGACGAAGA	GGGCCACCGG	11520
	CCCCGTCTGG	CGCTCCAGGG	CCGCCAGAAC	GCACGCATAC	AGCGCCCGCC	ACAGGGTCGG	11580

	GTCCCCAGG	GGCTCCAGCG	GGGAGGCGGC	CGGGGCGGCG	CGGGGCGGCG	CGACGGCCCC	11640
	GCGGCCGAGA	CGTCGGGGGA	GCCGTAGAAG	TCCTGCAGGT	CGGACGAACC	AACGGACACC	11700
	TCCGCAAAGC	GCGCGCGCGC	CTCCCCCGCG	GCGTCGCGAC	AGACCAGATA	CAGCAGGGCG	11760
	TGGAGGCAGT	CGCGCGTGCG	CGGGGGCAGC	CATACCGCGT	ATAGGGTAAT	GGCGCTGACG	11820
5	CTCTCTCTCA	CCCAAACGAT	GCCGGGGGCT	TCCATGCCAC	GACGCCCCGG	GGTTGCCGTG	11880
	TATCGAACGA	GCGCGGCCCC	AGACTTATAG	GGTGTAAAG	TTCACCGCCC	CCTGCATCAT	11940
	GGGCCAGGCC	TCGGTGGGAA	GCTCCGACAG	AGCCGCCTCG	AGAATGATGT	CAGTGTGGG	12000
	CTGGGCGCCG	GAGGCGTGCG	TGCGCAAGCA	GCGCCCCAC	GCGGGCGCGC	GCAGCTTGAA	12060
	GCGCGCGCCC	GCAAACCTCC	GCTTATGGGC	CATCAGCAGC	GCGTACAGCT	GTCTGTGCGT	12120
10	CCGGCAGGCG	CTGTGGTCTGA	TGCGGTGGGC	GTCCAGCAGC	TCCACGATGG	CTCGCTTGGT	12180
	GAGGTTTTTA	ACGCGCCCCG	CCCCGGGAAA	CGTCTGCGTG	CTCTTGCCCA	GCTGCACCCC	12240
	GAACAGTTCG	CCCCAGATGA	TCTTGAACAG	CGACAGCGCG	TGCTCCGTCT	CGCTCACGGA	12300
	CCCGCGCGGG	GGGCGAGCCG	TCAGGGCGTC	GGCCACGCGC	TTAACCGCGT	CCTCCGACAG	12360
	CAAGGGGCCG	TCGGTCACGT	TACAGTGCC	CAGTTCGAAC	ACCAGCTGCA	TGTAGCGGTC	12420
15	GTAGTGGGGG	TTCAGCAGCT	CCAGCACGTC	CTCGGGGCTA	AAGGTTCGCC	CCGACCCCCC	12480
	GGCCATCGAG	TCCCACTGCA	GGCACGCGCG	CATGGTGCTG	CACAGACGGA	ACAGCTCCCA	12540
	GACGGGGGCG	ACGTTTAGGG	TGGGGTGTAG	GGCCACAAGC	TCCAGCTCTC	CGGCGGCGTT	12600
	GATCGTGGGG	ATGACGCCCC	TGGCGTAGTG	GTCTGTAAGC	CGCCGGAAGA	TGGCGCTGCT	12660
	ATGGGCGGCC	ATGGGGACGC	GAAGACAGGC	CTCCAGCAGC	ACCAGGTAGA	TGAACCGCGT	12720
20	GCGGCCGACC	AGGCTGTTGA	GGCCGCGCAT	GAGCGCGACC	ACCTCGGCCG	GCGCGACGTC	12780
	CGGCCGAGG	TACTTTTCGA	CGAAAAGGCC	CACCTCCTCC	GTCTCGGCGG	CCTGGGCCGA	12840
	CAGGGACGTG	TCGGGGTCCCT	GGCAGCGCAG	CTCCCGCAGA	TCCCGCTGGG	CCCTCAGGGC	12900
	ATCAAAATGT	ATCCCCCGCA	AAAACAGACA	AAAGTTCCTC	GGGGTCAGCG	CGGCGTCGTG	12960
	GCCCCAGAAC	CGCACGTGCA	TGCAGTTGAG	GGTCAGAAAGC	ATGTGGAGGA	TGTTAAGACT	13020
25	GTCCGCGAGG	CACGCCAGCG	TGCACCTCTC	GAAGTAGTGC	TTGTACCGGA	ATTTGCTGTA	13080
	GATGCGCGAC	CCCCGCGCCT	GCGCCGCGTC	GGCGTGCGAC	GCGTCGCGAC	GCCCTTTGAA	13140
	CCGGCGGCAC	AACAGGTTCG	TCACCTGGGA	AACTGTGCC	GGCCACTGCC	CGCTGGCGCT	13200
	CACCACGTGG	TTGAGCAGCA	TGGGCGTAAA	GACGGGCTCC	GAGCGCGCCC	CGGACCCGTC	13260
	CATGTAGATC	AGCAGCTCCC	CCTTGCGGAG	AGTCCGTACC	CGCCCCAGCG	ACTGGTACAC	13320
30	GGACACCATG	TCCGGCCCCG	AGTTCATGGG	TTTCACGTAG	GCGAACATGC	TGTCAAAGTG	13380
	CGGCGGATCG	AAGCTAAGGC	CCACCGTCAC	GACCGTTGTG	TAGATGACCA	CCCGGTACCG	13440
	GCCCCATGTG	GTCACGTGCG	CGGGCGGGGT	GAGCGAGTGG	AGCAGCAGCA	CGCGGTCCGT	13500
	AAACTGCCGG	CAGAACCTGG	CAACGACCTC	CGCGAAGGAG	ACCGTCGACG	AGAAGATGCA	13560
	GACGTTATCT	CCGCCGGCCA	GGCGCGCCTC	CAGCTCCCCG	AAGAAGGTGG	CGTCCGGGGG	13620
35	GGCGTCCGGG	GGGGGCGCCC	CGCCGCGCCG	CCCCGGCGGG	CGCAGGGCCG	CCTGCAGGAC	13680
	CTCGGGCCCC	AGGCGCGGGA	GAAACAGACA	ACGCGCGGCC	GAAAATCCGG	GCATGGCATA	13740
	CTCCCCGATG	ACCACGTGAA	CGTTCTTTTC	GCCCCGGAGG	CTGCACAGAA	AGTCCACCAG	13800
	CTGCGCGTTG	GCGGTGGCGT	CCATGGCGAT	GATCCGCGGG	CACGTGCGCA	GCAGGCGCAG	13860
	CATCAACGCG	TCGACGCGGC	CCAGCTGCTG	CATCGTCGGC	GAGTACAGTT	GGCCCAACGT	13920
40	CGACATGACT	TCGTCCAGGA	CGAGCACGTC	GTAGTTGTTC	AACAGGTTTC	GGCCACGCG	13980
	ATGAAGACTT	TCCACCTGCA	CGATGAGACG	GTGAAGGGG	CGGTCTGTCA	TGATGTAATT	14040
	GGTGGATGAG	AAGTAGGTGA	CGAAGTCGGG	CAACCTGAC	TCAGCGAACC	GCGTCGCCAG	14100
	GGTCTGAGTA	AAACTCCGAC	GACAGGAGAC	GACCAGCACA	CTCGTGTCG	GAGAGTGGAT	14160

	CGCTTCCCCC	AACCAGCGGA	TCAGCGCGGT	AGTTTTTCCC	GAGCCCATTG	GCGCGCGGAC	14220
	CACAGTTACG	CACCGGGCCG	TCGGGGCGCT	CGCGTCCGGG	AAGGTGACGG	GTCCGTGTTG	14280
	CTGCCGCTCG	ATCGTTGTTT	TCGGGTGGAC	CCGGGGAACC	CACTCGGCCA	AATCCCCCCC	14340
	GTAAAGCATC	CGCGCCAGCG	ATACACTCGA	CGTGTACTGC	TCGCACTCGT	CATCCCCGAT	14400
5	GGGACGCCGG	GCCCCCAGGG	GATCCCCCGA	GGCCGCGCCG	GGCGCCGACG	TCGCGCCCCG	14460
	GGCGCGGGCG	GCGTGGTGGG	TCTGGTGTGT	GCAGGTGGCG	ACGTTTCATCG	TCTCGGCCAT	14520
	CTGCGTCGTG	GGGCTCCTGG	TGCTGGCCTC	TGTGTTCCGG	GACAGGTTTC	CCTGCCTTTA	14580
	CGCCCCCGCG	ACCTCTTATG	CGGAGGCGAA	CGCCACGGTC	GAGGTGCGCG	GGGGTGTAGC	14640
	CGTCCCCCTC	CGGTTGGACA	CGCAGAGCCT	GCTGGCCACG	TACGCAATTA	CGTCTACGCT	14700
10	GTTGTGTCGG	GCGGCCGTGT	ACGCCGCGGT	GGGCGCGGTG	ACCTCGCGCT	ACGAGCGCGC	14760
	GCTGGATGCG	GCCCGTCGCC	TGGCGGCGGC	CCGTATGGCG	ATGCCACACG	CCACGCTAAT	14820
	CGCCCGAAAC	GTCTGCGCGT	GGCTGTTGCA	GATCACAGTC	CTGCTGCTGG	CCCACCGCAT	14880
	CAGCCAGCTG	GCCCACCTTA	TCTACGTCCT	GCACTTTGCG	TGCCTCGTGT	ATCTCGCGGC	14940
	CCATTTTTCG	ACCAGGGGGG	TCCTGAGCGG	GACGTACCTG	CGTCAGGTTT	ACGGCCTGAT	15000
15	TGACCCGGCG	CCGACGCACC	ATCGTATCGT	CGGTCCGGTG	CGGGCAGTAA	TGACAAACGC	15060
	CTTATTACTG	GGCACCTCC	TGTGCACGGC	CGCCGCGCGG	GTCTCGTTGA	ACACGATCGC	15120
	CGCCCTCAAC	TTCAACTTTT	CCGCCCGGAG	CATGCTCATC	TGCCTGACGA	CGCTGTTCCG	15180
	CCTGCTTGTC	GTGTGCTGT	TGTTGGTGGT	CGAGGGGGTG	CTGTGTCACT	ACGTGCGCGT	15240
	GTTGGTGGGC	CCCCACCTCG	GGGCCATCGC	CGCCACCGGC	ATCGTCGGCC	TGGCCTGCGA	15300
20	GCACTACCAC	ACCGGTGGCT	ACTACGTGGT	GGAGCAGCAG	TGGCCGGGGG	CCCAGACGGG	15360
	AGTCCGCGTC	GCCCTGGCGC	TCGTGCGCCG	CTTTGCCCTC	GCCATGGCCG	TGCTTCGGTG	15420
	CACGCGCGCC	TACCTGTATC	ACCGGCGACA	CCACACTAAA	TTTTTCGTGC	GCATGCGCGA	15480
	CACCCGCGAC	CGCGCCCAT	CGGCGCTTCG	ACGCGTACGC	AGCTCCATGC	GCGGTTCTAG	15540
	GCGTGGCGGG	CCGCCCGGAG	ACCCGGGCTA	CGCGGAAACC	CCCTACGCGA	GCGTGTCCCA	15600
25	CCACGCCGAG	ATCGACCGGT	ATGGGGATTC	CGACGGGGAC	CCGATCTACG	ACGAAGTGGC	15660
	CCCCGACCAC	GAGGCCGAGC	TCTACGCCCG	AGTGCAACGC	CCCGGGCCTG	TGCCCCGACG	15720
	CGAGCCCATT	TACGACACCG	TGGAGGGGTA	TGCGCCAAGG	TCCGCGGGGG	AGCCGGTGTA	15780
	CAGCACCGTT	CGGCGATGGT	AGCCGTTTCG	TTCTGTTTAA	TAAACCGACG	TTGTGCGTTT	15840
	CACCATACTT	CGGCGCGCGC	GTGTGTGTGT	TTTTTTTGTG	GTGTTTATTT	TCCCCCACCC	15900
30	CTTCCTTTTC	TTTCGGCCAC	CACCCCCCTC	CTCCCCGTA	CTATACAACA	AAAAATACCA	15960
	CACATACGAC	CAAATACGGA	CAATCATTTT	TGTCCTTTAT	CGCTGTCAGA	GAGTGGGGGC	16020
	GTGAGCGTGG	CAGGAGGGCG	GGCCACGTCG	GGGTCCCGCC	GTCTGGTGTG	ACGCGATGGG	16080
	GGGTCCGATG	CGCGCCGGTA	CTGGGGCCCC	GGCGCCCGGG	TGACCACGCG	CATGTCGGGG	16140
	GGCACGTAGA	AGTTACCTTC	TTCTTCGGAC	TCGATGTCCA	CGACGTCAA	TTCTGTTGGC	16200
35	GTCAGCGAGA	CGACCTCCCC	GCCGTCCGGT	ATGATGACGT	TGTGTGCGCA	GCAGCAGGGC	16260
	CGCGCCCCGG	AGAACGCGAG	GCCCATAACT	TGGCGAGCGT	ATCGTCGAAG	GCCAGGCGGC	16320
	TGTTTTCGCCG	GATGTCCCGG	TAGATCCCCG	GCTCGACGCG	GACGGGGGTG	ATGATCAGGG	16380
	CGATCGGAAC	GGCCTGGTCC	GGGAGGATCG	ATGCCCTTGG	GGGTCCGGGG	GCCCCGCCAC	16440
	GCCCCGCGGG	CGCTCCGCGG	CCGTCTTCCA	GGCGGAACGT	CACGCCCTCC	TCCGCGCCCC	16500
40	CGCGGTGCCT	GCCGAGGAAC	GTCACCAGGT	GCGGTTGCAG	GGGGCAGTCG	GGAAAGTGGC	16560
	TGTCGAGGAC	GTATCCCTGC	ACCAAGATCT	GTTTGAAGTT	CGGGTGGCGG	GGGTTGGCGA	16620
	AGATGGGCTC	GCGGCGAACC	AGCTCCCCGG	AGCTCCAGGC	CACGGGAGAG	ATGGTGGCAG	16680
	GCTCGAGGTC	GGGGACGCCA	AACAGAAGCA	CCTCCGAGAC	AACGCCGCTA	TTTAACCTCA	16740

	CCAGCGCCCG	ATCCGGGGCG	GAGCATCGCC	TTTTTTCGCC	GGCGGCGCGG	GAATCGAGCC	16800
	AGTCCCGGTC	TTGGGTGACG	AGCGCCTCCT	CCGGGCCCCG	GACGCGCCCG	GGCGCGAAGT	16860
	AGCGCACGCC	GGGGTTGGGG	ATGGACCGGA	TGAACGCCCC	GAACGCCTCC	GGCGATCGCC	16920
	GCGCCATCAG	GTCCTCGTAC	GCGGAGGCCG	CGGGGGCGCC	GGGTCCGCG	GGGTCAACG	16980
5	CGTACTTGGC	TCGGCACTTA	ACCTCGTAGA	AGGCCAGGGG	GGTCTGGGGG	GCGGGGGCCA	17040
	GGTAGCCGTG	AGGGTCCCTG	GGGCACACGA	GGATGTCCAG	GGACGCCCCC	ACCATGCCCCG	17100
	TGTGGCCGTC	CATGAGGACC	CCGCACGCGT	GCACGTCTTC	CTCGGCGAGG	TCCCCGGGTT	17160
	GGTGAAAGAC	GAAGCGCCCG	GCGTCGGCGT	CGTCGTTGAC	GCCCCGCTCC	GCGCGGCCCCA	17220
	CGCAGTAGCG	AAACAGCAGG	TTTCGGGGCG	TCGGCTCGTT	CACCCGCCCCG	AACATCACC	17280
10	CCGACGACTG	GGCGTCCAGC	CGCAGGCTGG	CGTTGTGGGT	GAGCCACTGG	GACGAGAAGC	17340
	ACGGACCCTG	CGCGCCCCAC	CGCAGCGTGG	AGGCGGTCGT	CAGGCCCCCG	CGAAGCAGGG	17400
	CCCAGAGCTG	GCACTCGGCC	TGGTTTTGCG	TCGCCGCCCTC	GTAAAATCCC	ATAAGCGGGC	17460
	GGGGGGCGAC	GGCTTCGGCG	GCGGACGGGG	GGGCGCGGCG	CGTCAGGCGC	CAGAGGTGCC	17520
	GGCCGAGCCC	GCGGTCCACC	ATGCCGGCCG	CCTCCAGCGA	CACGACGAGG	GAGCACAGAT	17580
15	AGTCCAGGCG	AGCCCACAGG	GGCCCCGATG	CCAGAGGGGA	GCGGACGCCG	CGCAGCAGGC	17640
	CGCGCAGGTG	GCCTCGAAC	GTTTCCGCCA	AGATATGGGG	GGGCAGTGCG	TGGGGGATCG	17700
	CCGACGCCGA	CCACATCGGG	TCGGGGTCCG	GGGGACCGGG	GCTGCAGTCC	GGGTCGATGG	17760
	CGTGTGCGCC	CCCCGGCGAG	AGGGGAATGT	CGGGGGTTGG	CGGGCCGGAT	GAGGCCTCAG	17820
	AGAGGGCCGG	GGACGCGGGC	CGGGCCTTTT	CGCCCCGGGC	CCCCCGCTCG	GGTTGCCCCAC	17880
20	GTGGGGGGCT	CTGGGGCCAA	TGGGAACCCG	GGGCCCCCGG	TGACGTGGGG	CGGGGTGGGG	17940
	CGGGGCGGGG	CCCAAAGACG	GTCGCCAGAT	CTAGGCTGTT	GGGTCGGGGC	CGCTTCGGGG	18000
	GACTATCGGG	GTCGCGGGCG	GGGTCCGCGG	GGCGCTTGCG	GCCGGGTGTT	GCGGCGGGCCG	18060
	CCATTTTAC	GAGCAGCCGA	AGAGCTCGAG	GGCGGAAGGG	ATCCTCACGA	CAGAGAGTGG	18120
	CGCGCGGCCG	GGTTGGCGTG	ACAGAGGCGG	GAGACCAGCA	CCAGCAGCGG	CCTCAGCTCG	18180
25	GGCGGCAGCG	ACACCGACGA	CAGGACGGCC	TTGTGCGTGC	GCTGGTAATT	TATACACTGC	18240
	TCCGTGAACG	CGCGCCGAAT	CTTGGGATTG	CGAAGGTGGC	GCCGGATGCC	CTCCGGCACG	18300
	TCATACGCCA	GGCCGTGGGT	GTTGGTCTCG	GCCGAGTTGA	CAAAGAGGGC	GGGGTGCAGA	18360
	ACGCAGCGAT	AGGCGAGGAG	GGCCACGGCA	AAGTCCGGCG	AGAGCTGGTT	GTTAAAGTAC	18420
	TGGTAGCCCC	GGACGCGGGT	CACGGGGACG	CCCAGGCTCG	GGGCCACGTA	CACGCTAACC	18480
30	AGCAGCTCCA	GCAGCGTCTG	CCCCAGGGCG	TAGAGATCGA	CCGCCAGCCC	GACGTCGTGC	18540
	TTCAGGGGGC	GGTTGTTAAA	CTCGGCCCGC	TCGTTGTTGA	GGTACTTTAC	CAAGAGCTCC	18600
	GGCGGCTGGT	TGTACCCGTG	CCCCACCAGA	GTGTGAAAGT	TGGCCGTGGT	CAGGGCGGGC	18660
	GGCATCCCAA	ACCCCCGGGG	GGACTCGAGG	TCCGGCTCCT	GGAGGCAAAA	CTGGCCCCCG	18720
	GATATCGTGG	AGTTGGAGTT	CAGGGTCACC	AGGCTAAAGT	CGGCCAGGAC	GGCCCCCGG	18780
35	AGCGACACCG	CGTCCGATCG	CAGCATCAGC	AGGACGTTGG	CGCACTTGAT	GTCCAGGTGG	18840
	CTGATCCCGC	ACCTGGTGT	CAGGAACACC	ACGGCGCGCG	CCAGGTCTGT	GAAGCAGTGG	18900
	TGGAGGGCCG	TCGCGACGGA	GGGGTGGTC	GCGCGCAGGG	ACGCCAGCTG	GCCGATGTAC	18960
	TTGCCGAGGT	CCATGTCGTA	CGCGGGGAAC	ACGATCTGGC	GCTGCTGCAG	CGAGAACCCG	19020
	AGCGGGGTGA	TAAAGCCGCG	GATGTCGTGG	GTGCGGCCCG	CGCGAAGAGC	GCACTCCCCC	19080
40	ACGAGCAGGG	TCGCGACGAG	CTCCACGGCA	AACCACTCCT	TTTCCCGGAT	GGTCTTCACG	19140
	GCGAGCTTGT	GTTGCGGAAT	CAACTGCACC	TCGCCGTACC	CCCCCGAGCC	CCCGAAGCTG	19200
	CGGGCCCCGG	GGATCTCCAG	GGTCGTGTAG	CGGAGGGCGG	GGTTGACGGC	GAATACGGGG	19260
	ATGCATAGCT	TGTGGATGCG	CGCGAGGGAC	AGGATGTGCG	AGGGGGGCGA	CGGGGGCGAG	19320

	GTCATGGCCG	TCTCGGACCT	GCGCAGGGGC	GGGCGCCTCA	GCTTGGCCGC	AGGGCCGGGG	19380
	GCCTCGGGGG	ACGAGCGGCG	ACGAGACGAG	CGGCTCACTC	GCCATCGGGA	CAGTCCCGCG	19440
	CGAAGCCGCT	CCCGGAAGCT	GGATCGGCGG	CGGGACCCGG	GGCGGGCTCC	GGAGACGGCG	19500
	CCGTCTCGGG	GGGAGGGGCC	GCTTGGGCGT	CCGGACGCCC	GGCGGCTGAG	GGAGTGTATG	19560
5	TAGGACGCGA	GCCAGGCCTT	GAAGGAGCGT	CGGTGTGCAC	CTTGGGGGCT	GATGTCAGCT	19620
	GCCACATGAC	TAGCAGGTCG	CTGTGCCCCG	GACTCATCCA	TCCGTCCGCC	AGGTCGCCGT	19680
	CCCCCACAG	AGACGCGTTC	GCCGCGGCCT	CTTCGAGCTG	CTCCTCCTGG	TCCGCAAGAC	19740
	GATCGTCCGC	CGCGTCCAGG	CGCTCGCTAA	GCGCGGGATC	GAGGTACCGT	CGGTGTGCGG	19800
	TTAGAAAATC	ACGTGCGGCC	GCTTGCTCTT	CCACGCGAAT	TTTAACACAG	GTCGCTCGCT	19860
10	GTCGCATCAT	CTCTAAGCGC	GCGCGGGACT	TTAGCCGCGC	CTCCAATTCC	AAGTGGGCCG	19920
	CCTTGGCGGC	CATAAAGGCG	CCAACAAACC	TAGGATCTTG	TGTACTCACG	CCCTCCCGGT	19980
	GTAGCTGCAG	GGTCTGGTCC	CTGTACACCT	CGGCCCCGAG	GTGCGTCTCG	GCCAAACGTC	20040
	GGCGCAGGGC	CGCGTGGCTG	GCGTCTCGGC	TCATCTCGCC	GCCCCCGCGC	GCGCCCGACG	20100
	TCGGACTCCT	TCGCCCCGAC	CCCCCTGACC	TCAGCCGCCC	CCGCCTCGCC	CGCGATGTTT	20160
15	GGCCAGCAGC	TGGCGTCCGA	CGTGCAGCAG	TACCTGGAGC	GCCTGGAGAA	ACAGAGGCAA	20220
	CAGAAAGTGG	GCGTCGACGA	GGCGTCGGCG	GGCCTGACGC	TCGGCGGCGA	TGCGCTGCGC	20280
	GTCCCTTTTT	TGGATTTTGC	CACCGCGACG	CCCAAGCGCC	ACCAGACCGT	GGTCCCAGGC	20340
	GTCGGGACGC	TCCACGACTG	CTGCGAGCAC	TCGCCGCTCT	TCTCGGCCGT	CGCGCGGCGG	20400
	TTGTGTTTTA	ATAGCCTGGT	GCCGGCGCAA	CTCAGGGGGC	GTGACTTTGG	GGGCGACCAC	20460
20	ACGGCCAAGC	TGGAGTTCCCT	GGCCCCGAG	CTGGTGCGGG	CGGTGGCGCG	CCTGCGGTTT	20520
	CGGGAGTGCG	CGCCGGAGGA	CGCCGTGCCC	CAACGCAACG	CCTACTACAG	CGTCCTGAAC	20580
	ACGTTTCAGG	CCCTGCACCG	CTCCGAAGCC	TTTCGGCAGT	TGGTTCACTT	CGTGCGGGAC	20640
	TTCGCCCAGT	TGTTGAAAAC	CTCGTTCCGG	GCCTCTAGTC	TCGCGGAGAC	TACGGGCCCC	20700
	CCGAAGAAAC	GGGCCAAGGT	GGACGTGGCC	ACCCACGGGC	AGACGTACGG	CACCTTGGAG	20760
25	CTCTTCCAGA	AAATGATACT	AATGCACGCG	ACCTACTTTC	TGGCCGCCGT	GCTGCTCGGG	20820
	GACCACGCGG	AGCAGGTCAA	CACGTTCCCTG	CGGCTCGTGT	TCGAGATCCC	CCTGTTTAGC	20880
	GACACGGCCG	TGCGGCACCT	CCGCCAGCGC	GCCACCGTGT	TTCTAGTCCC	CAGGCGCCAC	20940
	GGAAAGACCT	GGTTTTTGGT	GCCCCTCATC	GCCTGTGCGC	TCGCGTCCCT	CCGGGGGATC	21000
	AAGATAGGCT	ACACGGCCCA	CATCCGCAAG	GCGACCGAGC	CCGTGTTTGA	TGAGATCGAC	21060
30	GCCTGCCTGC	GGGGCTGGTT	TGGCTCGTCC	CGGGTGGACC	ACGTCAAGGG	GGAAACCATC	21120
	TCGTTCTCGT	TCCCGGACGG	CTCGCGCAGC	ACGATCGTGT	TTGCCTCCAG	CCACAACACG	21180
	AACGTAAGTA	CGCCTTCCTC	CCGCGGTGCC	TGTTTCCCCG	GTGCCGCCCT	CCCCGAGATC	21240
	GACCGACAGA	CAAACACAGC	CAGACGCGAG	TGTGGGACGA	CACGCCCCGA	GCCCCCCCCG	21300
	CCATGGCGGG	GGGAAGCCTT	ACTGTTTATT	TGTAATCGGA	CGATGAGGCT	CTGGCCACGG	21360
35	CCCCGCGGAC	CGCGGGGCGAG	CTCGTTGCAA	ACAGGCGGCT	GGTATACGAT	GACAGAACGC	21420
	AGAGGCGCCA	CCCGGCGCTG	GTCGGGCGGA	TGACGCTTTC	CGCGCCGTCC	CGGCCACGA	21480
	CGACCTCGTG	CAGGTGGGCC	GTGATGCGCG	GGCGGCGGGT	CGCCTGCCGC	AGGATAACCG	21540
	CGTCCACGGG	GTGCCCCAAG	AGGAGCTGAC	ACAGGCTCGC	GTCCCCCGG	ACGGCCAGGG	21600
	TGCGCTGGGC	CATATTGGAC	CACATGCACG	GGGCGACGCA	GGGACAGGCC	TCCGCCACGG	21660
40	CGGGGGCGCG	CCACAGCGCG	TTGGCGGAAT	CGATGTGGGC	CGTCGGGGCG	CAGGCGCCGC	21720
	CTCCTCCCGG	GGGGTCGGTA	ATCCTGGATA	GCAGCCATCC	TAAATGGCGG	GCCCCGCTGC	21780
	CCGGGGGACA	GAGCGACCCC	AGGTATCAT	CCATGGCCCA	GCAGTATATG	CGGCCGCGCG	21840
	GGAGGTGCCA	CCAGGCCCCC	GGACCCAGGG	CACAGCACGC	CCCCGGATTC	GGGGGCGGTT	21900

	CCGTGGGTAC	CAGGTAGGCG	CCGTCGAGCT	CGTGGGCCAC	GGGCTCGTCC	GCGAGCTGTT	21960
	CGGCGGCGGG	GTCGGGGGTT	TCCTCCGGGG	GGGAGGCAGC	TTCCAGGTGG	CCGAAGGCTA	22020
	GGGTGCACAG	CAGCGGGGTC	CGGGGGTGCG	TTACGCTGCG	GAGGTGGACG	GTGGCGCAGT	22080
	AGCGGCGCTC	GCGGTAAAG	AAGAAAATGG	CAAAGAACGT	GTTCGAAGGC	AGGCGCAGCG	22140
5	CCTTGGGCCG	CGTCAGGTAC	AGGAAGATCT	CGCAGAAAAG	GGCACGCTCG	GGGTCGGGGT	22200
	CCGGAAGGGC	CACCTGGCAC	AGCGGCTCGG	TGAGGACCGT	GAGGCACCGA	AAAATCTTAA	22260
	GCCGCTCGTC	CCCCCGAACG	ACGCGCCACA	CGAAGACAGA	GTTGGCGATG	CGCGCGACGA	22320
	GGTCGGCTTC	GGGCCCCGGG	TCGGGGGGCG	GCGCGTCGGG	GGGGGCGCCC	CGGTGACCCG	22380
	GCGGGGCGGC	GGCTCCCGGG	GGGCCTGGCG	TCGCCTGGGG	ACGCCAGAGT	GCCCCGTGTG	22440
10	CCAGGTTGGT	GGTGGGGAAG	GGACCGGAGA	CGCACCAAAA	GCAGAGGGGC	CAGCGCGTGT	22500
	ATGAGTTGGG	GGGGGGGTGG	GTGAGCGGTG	GAACAAAAGC	ACGCGTCAGC	GGACAAGGCC	22560
	GGGTCCCGTA	GCCGCCCGGC	GACAGAACCG	GAGTCCGACG	GCACGCGCGA	CGGGGTCTGC	22620
	GAGGCTGAGG	TACGCCGCGG	TGTTAATGGT	AAACGCAAAG	CCTCCCGGAA	AGACCACTAG	22680
	CCCGCAGAGG	CGGCGATTGA	ACCCAAGGCA	GAGGTACCGG	TAGCTCTCTC	CCGGAAGGTA	22740
15	TTGCTCGCAG	ACCCTGTGCG	GGGCAGTGGA	GGGGCTGCCC	TCCATGAAGC	GACATTTACT	22800
	CTGCTCGCGT	CCATTGACGT	CACCGTCAAT	CACCACTGCG	ATTGGACGGT	TGGTAAGGCG	22860
	CAGCGTGTCT	CCGCTGGTGC	TGTAGTAGTC	AAACGCGTAG	TGGGCGTCGG	AGTCGGCGAA	22920
	GCGGGCGGGG	ATGTCGTCGC	TGAGAGGGAC	GAGCCGCGCG	CGCCGCCCCC	GACCGCCCTG	22980
	GCCGCCCAGA	TGCGCCAGCA	CGGCCAGGGC	GTACGCGGTG	TGAAAGAACG	CGTCGGGGGC	23040
20	GGTCCCTCTG	AGGGCGCGCA	TCAGGTTCTC	CAGGAGCACG	GGGAAGCGCC	GCGTCACCTC	23100
	CCCTAGCCAC	TCGCTCTGGT	GGGGGCCAAA	GTCGTAGCGC	AGGCGCTGGA	AGATGCGCGG	23160
	GCCGCTTGG	AGCGCGGCCC	GGATAGAGTG	GCCCAGGGCC	CGCAGACACG	CGATCTGGAT	23220
	GCGCGCGACG	AAGGCCACCT	CGGCCGCGAT	GTCAAAGGGC	TGCAGCACGG	GGCGCGGGTG	23280
	GCGCAGGGGT	CCCTCGAGCG	CGGGAAGCGG	ACGCAGCAGC	GCCGTCTGGG	CCGCGGGGGA	23340
25	CAGCTGGTGG	GGGCGCACGA	CGCGCTCGGC	GGCACAGGCC	TCCGTCAGGG	CCGTGGCCAG	23400
	CTCGGAGGAC	AGCCGCGGGG	GGCGGGCGCG	TCGCCCCGCC	CACGCCACCG	AATTCTCGTA	23460
	GGAGACGACG	ACGAAGCGCT	GCTTGGTCCC	GTAGTGATGG	CGCAGGACCA	CGGAGATGGA	23520
	GCGACGGCTC	CACAGCCAGT	CGGGCCGGTC	GCCGCCGGCC	AGAGCTTCCC	ACCCGCGGTC	23580
	CAGCCACTCG	ACCAGCGATC	GCGGCTTGGC	GGTCCCCGGC	ACGAGGGTGA	GCACGTCGTT	23640
30	GAGGACGTCC	TCGCCGCGG	CCCGGGGGCC	CCCCGGGGTG	GCAAAGCGCC	CCCCGCGGGG	23700
	CGGCTCCAGG	CCCGCCAGCA	CCGCTCCGCG	GTCCGACGCG	CCCAGGGCTC	CCCCGCTGAC	23760
	GGCCTGGTGG	ACCAGGGCGC	CCTGGCGGAG	CCCCGAGGCG	ACGCCGGAGG	CCGCGTGCTT	23820
	GGGGCGCGCG	CGGACCGGGT	GGCGGCGGGT	GACGTCCTGC	ACGGCCCGCT	GGACCAGCGC	23880
	GAGGATCTCC	TCGTCTCTTT	GCGTGATGGA	CACGTCCTCC	GCGGTGGCCG	TGTCGCCTCC	23940
35	CGGGGCCGTG	AGCTGCTCCT	CCGGGGAGAT	GGGGGGGTCT	GGGGTGCCGA	CAACGGCCCG	24000
	CCCGGCCCCG	CCCGAGACCG	AGGACGCCCT	GGGAGTGGGG	GTGCCGCTTT	CCCCCATCCC	24060
	CAGGGACAGG	TGGGCCGCGG	CCTCCGTCGC	GGCGGCGGGA	GCCGCGGCCC	CCAGCCGCGC	24120
	GACGTAGCGA	CAAAAGTGGC	GACAGAGGCG	CATGAGGCGC	GCGCCGTCGG	CCGCGTATCG	24180
	CGTGTTTGGC	GGGACGAGCT	CGTCGTAAC	GAACAGGAGC	ACGCGGGCGC	AGGTCGCCCCA	24240
40	CGGGCCCCAC	GCCAGGCGCA	GCGCCGCGAC	CGTGTAACGG	TCGTACACGC	CTTGGGCGTC	24300
	GCACGCGACC	GGCAGGGAGA	CGAACAGCCC	GCCCCGCGTG	GGGACGCGCG	GCAGGAGGTC	24360
	CGGGTGCGCC	GGGATGACGG	GGGCTAGGAT	CGCCCCCACC	GCATCCGCGG	GCACGTAGGC	24420
	GGCAAACGCC	GAACGCCACG	GGGTGCAGTC	GCCGGTCGCG	TGGGCCCCGG	TCTGGGTTC	24480

	GACCCGGAAG	TTCGCGGCCG	CCCCGCCGTC	GGGGCGGCCG	CGCACGAGGG	CGGACAGCGG	24540
	GACCCCCGCC	GCCGCCAGGC	ACTCGCTGGA	GATGATGACG	TGAATCAGCG	AGGCGGGGCT	24600
	GCTCGGGTCC	CGGGTGAGAT	CGTATGGAC	CTCGTTGGCA	AAGTGCCCGT	TCATGGCCCCG	24660
	GCCGGCGGTG	CGAGCCCTTC	CCGGTGCCGG	AAGGGGCGTG	GGTGGGGGGT	GCGTGTGCGC	24720
5	GTCTTCGGGG	CCCCGCGGCG	CACGTGCGCT	TATACGCTGT	GTGTTCGTC	TGTCCCCAGG	24780
	GAATCCGGGG	CCAGGACTTT	AACCTGCTTT	TCGTGACGA	GGCCAACCTT	ATTCGCCCCG	24840
	ATGCGGTCCA	GACGATTATG	GGCTTCTCTA	ATCAGGCCAA	CTGCAAGATC	ATCTTCGTCT	24900
	CGTCGACCAA	CACCGGAAG	GCCAGCACGA	GCTTTTGTGA	CAACCTCCGC	GGGGCCGCCG	24960
	ACGAGCTGCT	CAACGTGGTC	ACCTATATAT	GCGACGACCA	CATGCCCGCG	GTGGTGACGC	25020
10	ACACCAACGC	CACGGCCTGT	TCCTGCTATA	TCCTGAACAA	ACCCGTGTTT	ATCACGATGG	25080
	ACGGCGCCGT	TCGCCGACG	GCCGATCTGT	TTCTGCCCGA	CTCCTTCATG	CAGGAGATCA	25140
	TCGGGGGGCA	GGCCCCGAG	ACCGGCGACG	ACCGGCCCGT	CCTAACAAAG	TCGGCGGGGG	25200
	AGCGGTTTCT	GCTGTACCGC	CCCTCCACCA	CCACCAACAG	CGGCCTGATG	GCCCCGAGC	25260
	TGTACGTGTA	CGTGGACCCG	GCGTTCACGG	CCAACACGCG	CGCTCCGGC	ACCGGCATCG	25320
15	CGGTCGTGCG	GAGGTACCGC	GACGATTTC	TTATCTTCGC	CCTGGAGCAC	TTTTTCCTCC	25380
	GCGCGCTCAC	GGGATCGGCC	CCCGCGGACA	TCGCCCGCTG	CGTCGTGCAC	AGCCTCGCCC	25440
	AGGTGCTGGC	GCTGCACCCC	GGGGCGTTTC	GCAGCGTTTC	CGTGGCGGTC	GAGGGCAACA	25500
	GCAGCCAGGA	CTCGGCCGTG	GCCATCGCCA	CACACGTGCA	TACCGAGATG	CACCGCATCC	25560
	TGGCCTCGGC	GGGGGCCAAC	GGCCCCGGGC	CCGAGCTCCT	CTTCTATCAC	TGCGAGCCGC	25620
20	CCGGCGGCGC	GGTATTGTAC	CCCTTCTTTC	TGCTCAACAA	ACAGAAGACG	CCCGCCTTCG	25680
	AATACTTTAT	CAAAAAGTTC	AATCCGGGG	GCGTCATGGC	GTCCCAGGAG	CTCGTCTCCG	25740
	TGACGGTGCG	CCTGCAGACC	GACCCGGTCG	AGTATCTGTC	CGAGCAGCTC	AACAACCTCA	25800
	TCGAAACCGT	CTCTCCCAAC	ACCGACGTCC	GCATGTACTC	CGGAAAACGC	AACGGTGCCG	25860
	CGGACGACCT	CATGGTCGCG	GTCATCATGG	CCATTTACCT	GGCGGCCCCG	ACCGGGATCC	25920
25	CCCCGGCCTT	TTTTCCGATC	ACGCGCACGT	CTTGAGTCTT	TCTTGCCGTT	TCTTTGTGTT	25980
	CTCTTTCTTT	CCCCCTCTC	TCCGCAATAA	ACGCCTTCCC	GGAAGTGTGT	TTTCCCCCCC	26040
	TACAACAGTG	TTGTCCGTTG	GTTGGGTGGT	TGGGGTGCGG	GGGTGGGCGG	GGGAAGCAAG	26100
	AAAACGGTCG	GCGAACACAA	CATCGGGAAA	ACGGATTCCC	GAACGTGCGT	CTTCCCAGAT	26160
	TCGACACACA	CCCCCTTCT	CCTTAAATAA	ACACAAACCA	CACGCTCGTT	GGTTGGTTAA	26220
30	TGCCGGCGCT	TTATTTACGT	CTTGTTTTTT	TGCGTTTCTT	CCGCGGGTCC	CTTCCCAACA	26280
	CGCCTGCCCC	CGCCTCAGGG	GTAGCGGATA	ACCGGGGCCA	TGTCGCCGGA	TTGCACAACG	26340
	GCGGCGCCGT	CGAACGTACA	CACCCGAACC	GCCGGGGCCA	GGGCCAGGAT	GTCCCCGAGT	26400
	TGGCCCCGCT	GCGCCAGCCA	GCGCACCAGC	GCCTCGTAAA	GCGGCAGCCT	GCGCTCGCCG	26460
	TCCTGCATCA	GCATGGGGGG	TTCGGGGTGG	ATGAGCTGGG	CGGCTTCTCG	CGTGACGCTC	26520
35	TGCATCTGCA	GGAGCGCGTT	CACGTATCCG	TCCTGGGCGC	TCAGCGCGAG	CAGCCGGGGG	26580
	ATGAGCGTGA	GGATGAGGGT	GGTTCCTTCG	GTTATGGAGT	AGACCATGTT	GAGGACGAGC	26640
	GACCGCAGCT	CGGTGTTTAC	GGAGGCGAGT	TGCTGGACGT	CGGCCACGAG	CGAGAGACGG	26700
	GCCCCGTTGT	AATACAGCAC	GTTGAGGTCG	GGGAGCTCCC	CGGGCGTCCG	GGGGTCGGGG	26760
	TTGAGGTCCC	GGATGCCCCG	GCGCACCAGC	CGCGGACTA	TCTCGCGGGC	CAGGGGCGTT	26820
40	GGGAGCGGGA	CCGGAAACCG	CAGCGTGAGG	TCCAGCGACT	CCAGGCGCAC	GTCCGTCGCC	26880
	TGGCCCTCGA	AGACGGGCGG	GACGAGGCTG	ACGGGATCCC	CGTTGCAGAG	GTCGACGGGG	26940
	GAGGTGTTGC	GGAGATTGAC	GGTGCCGGCG	TGCGTGAGCC	CCAGGTCCAC	GGGGCAGGCG	27000
	ACGATTGCGG	TGGGCAGCAC	CCGCGTGATT	ACCGCGGGGA	AGCGCCTGCG	GTACGCCAGC	27060

	AACAACCCCA	ACGTGTCGGG	ACTAACTCCT	CCGGAGACGA	ACGATTCTGT	CGCCACGTCC	27120
	GCGAGCGCCA	GCTGGCGGCG	GATGGTCGGC	AGAAAGACCA	CTCGACCCTC	GCACCGCTGC	27180
	AGCGCCGCGG	CATCGGGGCG	CGAGATACCC	GAGGGGATCG	CGATGTCTGC	TTCGAAACAA	27240
	TCCGTGATCA	TGGCGCCGGG	CCGCGAGACA	CCGGAACGCG	GGGGTGCGGG	AGGGCCGGAA	27300
5	AGCGCAACGC	AACCGGGACG	ATGATGAAAC	AGAGATGGGG	GGCACCAGCC	GTGTGGGAGA	27360
	GGGGGCGGGG	CAGGGCTCAG	CAGCACGCAC	GGGGAGGTCT	GTCGTGCGCA	GGAGCCCCAG	27420
	GTGAGAATCA	GTCCCCCGGA	GCTCGGGTCT	GGGTTTATT	GGGACCTGCC	CTCGGAATCG	27480
	CGGCTCCCAG	TCCAAGCCCC	CCCGGGGGGG	CGGGGACAGG	GGGTGTGTGT	GGGTAAAAGC	27540
	AACGTCCGAA	AATCAAACCC	AATGCCCCAA	ACAGGAAAAA	AAAAAAGAC	GGGCGGGTGG	27600
10	AGGGAAAGCT	GGGGAAGAAG	AAGCCAATTT	TACAGAGACA	GGCCCTTTAG	CGGGGAGGCG	27660
	TCGTAGATGA	GATACTGCGT	AAAGTGGGTC	TCTCGCGCGT	GGGCCTCCCC	ATCGCGGGCG	27720
	CTGCGTAGCA	GGGCGGGGTC	GCTGGCGCAG	GTGATCGGGT	AGGCTTCTTG	AAACAGGCCG	27780
	CACGGGTCTT	CCACGAGCTC	GCGGCACCCC	GGCGGGCGCT	TAAACTGCAC	GTCGCTGGCA	27840
	GCGGTGGCCG	TGGATACCGC	CGATCCCGTT	TCCACGATAA	GACGCTCCAG	GCAGCGATGT	27900
15	TTGGCCGTGA	TGTCGGCCGC	GGTGAAGAAC	TTGAAGCAGG	GGCTGAGGAC	GGGCGAGGCC	27960
	CCGTGAGGT	GATAGGCCCC	GTTGTACAGC	AGGTCCCCGT	ACGAGAACCG	CTGCGACGCC	28020
	CACGGGTGG	CCGTGGCCGC	GAAGGGCCGC	GCCGGGTGCG	TCTGGCCGTG	GTCGTACATG	28080
	AGGGCTATGA	CGTCCCCCTC	CTTGTCCTCC	GCGTACACGC	CGCCGGCCGC	GCGTCCCCGC	28140
	GGGTGTCAGG	GCCGGCGAAA	GTAGTTGATG	TCCGTGGCCA	CGGGGGTGGC	GATGAACTCA	28200
20	CACACGGCAT	CCTGCCCCGT	GTCCATGCCG	GCGCGCCGCG	GCACCTGGGC	GCAGCCAAAG	28260
	ACCGGGAGGG	GCTGGGCCGG	CCCCAGCCGG	TTTCCCGCCA	CGACCGCGTT	GCGCAGGTAC	28320
	ACGCGGCCCG	CGTTGTCTAG	CAGCGGGGGG	GCCCCGCGGC	CGAGGTAAAA	GTTTTGGGGG	28380
	AGGTGCCCCA	TGTCCGTAAC	GGGGTTGCGG	ACGGTGGCCG	TGGCCGCGAC	GGCGGTGTAG	28440
	CCCACACCCA	GGTCCACGTT	TCCGCGCGGC	TGGGTGAGCG	TGAAGCTGAC	CCCCCGCCCC	28500
25	GTTTCGTGGC	GGGCCACCTG	GAGCTGGCCC	AGAAAGTACG	CCTCCGACGC	GCGCTCGGAA	28560
	AACAGCACGT	TCTCGGTCAC	GAAGCGGTCC	TGCCGCACGA	CGGTGAACCC	GAACCCGGGG	28620
	TGGAGGCCCG	TCTTGAGCTG	GTGATACAGG	GCCACGGGGC	TCATCTTGAA	GTACCCCGCC	28680
	ATGAGCGCGT	AGGTCAGCGC	GTTCTCCCCC	GCCGCGCTCT	CGCGGGCGTG	CTGCACCACG	28740
	GGCTGGCGGA	TGGAGGAGAA	GTAGTTGGCC	CCCAGGGCCG	GGGGGACCAG	GGGGACGTGG	28800
30	CGCGCCAGGT	CGCGCAGGGC	CGGGGGGAAG	TTGGGCGCGT	TGGCCACGTG	GTCGGCGCCC	28860
	GCAAACAGCG	CGTGGACGGG	CAGGACGTAG	AAGTATTCGC	CATTTTGGAT	GGTGTGGTCC	28920
	AGGTGCTGGG	GGGCCATGAG	CAGCACGCCG	GCGTGCAGCG	CCCCGTGCGA	GATGCGCATG	28980
	TTGGCCGTCG	ACGCGGTGTT	GGCGCCCGCG	TCGGGCGCCG	CGGAGCACAG	CAGCGCCGTC	29040
	GTGCGCTCGG	CCATGTTGTG	CGCCAGCACC	TGCAGCGTGA	GCATGGCGGG	CCCGTCGACG	29100
35	ACGACGCGCC	CGTTGTGGAA	CATGGCGTTG	ACCGTGTTGG	CCACCAGATT	GGCGGGATGC	29160
	AGCGGGTGGG	CGGGGTGCGT	CACGGGATCG	CTCGGGCACT	CCTCGCCGGG	GGCGATCTCC	29220
	GGGACCACCA	TGTTCTGCAG	CGTGGCGTAC	ACGCGGTGCA	AGCGGACCCC	CGCGGTGCAG	29280
	CAGCGCCCCC	GCGAGAAGGC	CGGCACCAGC	ACGTAATAGT	AGATTTTGTG	GTGGACGGTC	29340
	CAGTCGGCCG	GCCGGTGCAG	CCGGTCGTCG	GCGGCGTCGG	CCGCGCGGGC	CTGGGTGTTG	29400
40	TGCAGCAGCC	GGCCGTCGTT	GCGGTTAAAG	TCGGCCGTCG	CCACGTTGCA	CGCCGCCGCG	29460
	TAGACGGGCT	CGTGTCCCCC	CGCGTCAATC	CGGCAGTCTC	GGTGGCGGTC	CAGGGCCGCG	29520
	TGTCGCATAA	GGCCGTCGCA	GTCCACACAG	AGGGGCGGCA	GCAGCGCCGG	GTCGCGCATC	29580
	AGGTGATTCA	GCTCGGCCCT	AGCCTGCCCC	CCCAGCTCCG	GGCCCGGCAG	GGTAAAGTCG	29640



	TCCACCAGCT	GGGCCAGGGC	CTCGACGTGG	GCCACCAGGT	CCCGATACAC	GGCCATGCAC	29700
	TCCTCGGGGA	GGTCGCCCCC	GAGGTAGGTC	ACGATGTACG	AGACCAGCGA	GTAGTCGTTC	29760
	ACGAACGCCG	CGCATCGCGT	GTTGTTCAG	TAGTGGTGA	TGCACTGAGT	CACGAGCCGC	29820
	GCCAGGGCGC	AGAACACGTG	CTCGCTGCCG	TGAATCGCGG	CTTGACAGCAG	GTAAACACC	29880
5	GCCGGGTAGC	TGCGGTCCTC	GAACGCCCCG	CGGACGGCGG	CTATGGTAGC	CGGCGCCATG	29940
	GCGTGGCGGC	CAACGCCGAG	CTCCAGGCCC	CGGGCGTCAC	GAAACGCCAC	CGGACACAGC	30000
	GCCAGGGGCA	GGTTGCCGTT	GACCACGCGC	CAGGTGGCCT	GGATCGCCCC	CGGACCGGCC	30060
	GGGGGGACTT	CGCCGCCGGG	AAGCTCGACG	TCGGCCACGC	CCGCGAAGAA	GTCGAACCGC	30120
	GGGTGCAGCT	CCAGAGCCAG	GTTGGCGTTG	TCGGCTGCA	TGAACTGCTC	CGCGGTCATC	30180
10	TGGCACTCGG	CGACCCACCG	GACCCGGCCG	TGGGCGAGGC	GCTGCCGCCA	GGCGTTCAGA	30240
	AAACGCTGCT	GCATGTCCGC	GCCGGGGCCG	GCCGGGGCCG	CGACGTACGC	CCCGTACGGA	30300
	TTTCGCGCCT	CGACGGGGTC	GTGGTTCACG	CCCCGACGG	CCGCGTCGAT	GTTTCATGAGC	30360
	GAAGGATGAC	ACACGGTCCC	GACCGCGTTC	TCCATGGACA	GCCGCAGAAC	CTGGTGGTCC	30420
	TTTCCCCAAA	AAAACAGCTG	CCGGGGAGGG	AACGCGCGGG	GCTCCGGGTG	GCCGGGGGCG	30480
15	GGCACCAGGT	CCCCGGCGTG	CGCGGCGAAG	CGCTCCATGG	CCGGGTGAA	CAGCCCCAGG	30540
	GGCAGGACGA	ACGTCAGGTC	CATGGCGCCC	ACCAGGGGGT	AGGGCACGTT	GGTGGCGGCG	30600
	TAGATGCGCT	TCTCCAGGGC	CTCCAGGAAG	ACCAGCCTGT	CGCCTATGGC	CACCAGATCC	30660
	GCGCGCACGC	GCGTTGTCTG	GGGGGCGCTT	TCGAGTTCAT	CCAGCGTCTC	CCGGTTCGCC	30720
	TCGAGTTGCT	CCTCCTGCAT	CTCCAGCAGG	TGGCGGCCCA	CGTCGTCCAG	GCTCCGCACG	30780
20	GCCTTGCCCA	TCACCAGCGC	CGTGACGAGG	TTGGCCCCGT	TCAAGACCAT	CTCGCCGTAG	30840
	GTCACCGGCA	CGTCGGCCTC	GGTGTCTCTC	ACCTTCAGGA	AGGACTGCAG	GAGGCGCTGT	30900
	TTGATGGCGG	CGGTGGTGAC	CAGCACCCCG	TCGACCGGCC	GCCCCGCGCT	GTCGGCGTGC	30960
	GTCAGGCGGG	GCACGGCCAC	GGACGGCTGC	GTGCGCGTGG	TCAGGTCCAC	GAGCCAGGCC	31020
	TCGATGGCCT	CGCGGCGATG	GCCCGCCTTG	CCCAGGAAGA	AGCTCGTGTC	GCAAAAGCTC	31080
25	CGCTTCAGCT	CGGCGACCAG	GGTCGCCCCG	GCGACCCTGG	TCGCCAGGCG	CCCGTTGTCTG	31140
	AGATATCGTT	GCATGGGCAA	CAGCAGGGCC	AGGGGAGGCG	CCTTCTCCAA	CAGCACGTGC	31200
	AGCATCTGGT	CGGCCGTGCC	GCGCTCAAAC	GCCCCAGGA	CGGCCTGGAC	GTTGCGCGCG	31260
	AGCTGCTGGA	TGGCGCGCAG	CTGGCGATGC	AGGCTAATGC	CCGTCCCGTC	CAGGGCCTCC	31320
	CCCGTGAGCA	GGGCAATGGC	CTCGGTGGCC	AGGCTGAAGG	CGGCGTTCAG	GGCCCGGCGG	31380
30	TCGATGACCT	TCGTATGTA	ATTATGCACG	GGCTGCTCGA	CGGGGTGCGG	GCCGTGCGCG	31440
	GCGATGAGGG	GCTGGTGGAC	CTCGAACTGC	ACACGCCCTT	CGTTCATGTA	AGCCAGCTCC	31500
	GGGAACCTTG	TGCACACGCA	CGCCACGGAC	AGGCCGAGCT	CCAGAAAGCG	CACGAGCGAC	31560
	AGGGTGTGTC	AGTAGGACCC	CAGCAGGGCG	TCAAACCTCTA	CGTCATACAG	GCTGTTTTTCG	31620
	TCGGAGCGCA	CGCGGGCGAA	AAAATCAAAG	AGTCTGCGGT	GGGACGCCAC	CTCGATCGTA	31680
35	CTCAGGATGG	AGCCGGTGGG	CACCATGGCC	GCGGCGTACC	GGTAACCCGG	GGGGTTCGCGG	31740
	GCAGGAGCGG	CCATTGGGTT	CCTTGGGGGA	TTGCGAGGCT	CCATCAAGCC	AAGCTCGGGA	31800
	AGGCCAAGCC	CCTCCACAC	AACGCCCTAC	CGCCGGCGGA	CGGACTAAC	AACCCACGGG	31860
	CCGCCAAAAA	CCCCAAGGGG	CAACCCGACC	AACAACAGGC	GAGGGGAGGA	AAGGCGTAAA	31920
	GGGGGCGTTG	GGAGGCAAAA	AGAAAGAAAA	CACCCAGACG	TAGGCCCGAG	GACCGGCCGG	31980
40	CGTCTCTGT	CCCCGAGCAC	CCACTGTGCC	CAACAGGCAC	GGGGGCGAGC	TGCCCTTGCC	32040
	TTATATACCC	CCCCGCCACA	CCCCCGTTAG	AACGCGACGG	GTGCCTTCAA	GATGGCCCTG	32100
	GTCCAAAAGC	GTGCTAGAAA	AAAGTTGGTA	AAGGCGGCAA	AGCAGTCCGC	CGCCGCCACC	32160
	CACATGGCGG	CGCCGGCCGC	GCAGGCGATT	CCCAGAGAAC	GGGCGCGGAG	GGGATCCGTG	32220

	CGGGGCAGCA	GCTGGCTGGC	GGTGATCCAA	TGGAAAAGCC	CGTCGGGACT	GAACGTCTCA	32280
	TGGGCGGCCG	CCACCAGGGC	GCACAGGGCC	GCGCCGCCCA	TGATCACGCA	CAACCCCCAA	32340
	AACACGGGTG	GCGACAACGG	CAGGCGATCC	CGTTTGATGT	TCACGTACAG	GAGGAGCGCC	32400
	CGTGCCAGCC	ACGTGACATA	GTAGGCGAGG	ACGGCGGCTA	TAATACATGC	C GGCGCCACC	32460
5	GCCCCGTCCG	TCCACCCGTA	ATACATGCCC	GCGGCCACCA	GCTCCAGCGG	CTTGAGGACC	32520
	AGGAACGACC	AAGCAAACAT	CACCACCCGC	TTGGAAGA	CCGGCTGGGT	GTGGGGCGGA	32580
	AGACGCGAGT	AGGCCGAAC	GACAAAAA	TCAGACGTGC	CGTACGAGGA	CAGCGAAAC	32640
	TGTTTCATCGA	GCGGCAGTTC	GCCGTCTCTC	CCGCCACACG	CGGCCTCGTA	TACCAGCTCG	32700
	CGATCCAACA	AAGGAACATC	ATCCCGCATT	GTCTAGGTTCG	GTGCGGGGAG	CCGGCGAGGC	32760
10	AGCGAAACCG	AAAGTAGTGC	TGGCGGCGCG	GGCCCGGGTC	CGGACCCAAG	CTTCAGGGAT	32820
	GGGGGGCGGA	GGCCAAAATC	AAACAAGCAC	C GCGCGGGT	CTACACACAA	CCCCACCCG	32880
	GGTAGTATCC	GCGGATGCGA	GTGCCTGGCG	AAGTCACGTC	CCAGCAGGAT	ATAAACCTCG	32940
	GCCGTGGGTC	CCGGAACCCC	CGAAATTCAC	ACCCACGCCC	TGACGCCCAA	ATCATGGGTG	33000
	GATGTGGTTC	GCGAGCCGCA	CATCCGTGCG	TCCGCCCTCC	CCCGCGGGCT	GATGACGTGG	33060
15	CGGTTAGTCA	GTGGGAAGGC	AGGGGGAAAG	ATGGGTGGG	GGAGGAAACG	AAGAAAACAC	33120
	CCAGAGGGCC	ACGTCGGGAA	TGCGCCCGGA	GTTGTCTT	AAAGCCGCGC	CGTGCGTGAC	33180
	GGAAGCCGTC	GTTTGCCCAA	GCACCGACGC	C GCGATCCAC	AGTGGGGGGA	GTTCTCTCCG	33240
	CCGGCCACAA	CCCTACGCGC	GGGCGGCACG	C GCGAGAGCA	ACCCACGGGT	CCCGTTCGCG	33300
	CCACCGCCAG	CCCTTGCTCC	CACCACCTC	CTCCACCAC	CCCACTATTC	CCCCCCCCC	33360
20	AAGTCCGCCC	CGTGGCTCGC	CGGCCATGGA	GCTCAGCTAT	GCCACCACCC	TGCACCACCG	33420
	GGACGTTGTG	TTTACGTCA	CGGCAGACAG	AAACCGCGCC	TACTTTGTGT	GCGGGGGGTC	33480
	CGTTTATTCC	GTAGGGCGGC	CTCGGGATTC	TCAGCCGGGG	GAAATTGCCA	AGTTTGCCCT	33540
	GGTGTGTCGG	GGGACAGGCC	CCAAAGACCG	CATGGTCGCC	AACTACGTAC	GAAGCGAGCT	33600
	CCGCCAGCGC	GGCCTGCGGG	ACGTGCGGCC	CGTGGGGGAG	GACGAGGTGT	TCCTGGACAG	33660
25	CGTGTGTCTG	CTAAACCCGA	ACGTGAGCTC	CGAGCGAGAC	GTGATTAATA	CCAACGACGT	33720
	TGAAGTGCTG	GACGAATGCC	TGGCCGAATA	CTGCACCTCG	CTGCGAACCA	GCCCCGGGGT	33780
	GCTGTGTGAC	GGGGTGCGCG	TGCGCGCGCG	AGACAGGGTC	ATCGAGCTAT	TTGAGCACCC	33840
	GGCGATCGTC	AACATTTCCT	CGCGCTTCGC	GTACACCCCC	TCCCCCTACG	TATTCCGCCCT	33900
	GGCCCAGGCG	CACCTCCCCC	GGCTCCCGAG	CTCGCTGGAG	CCCCTGGTGA	GCGGCCTGTT	33960
30	TGACGGCATT	CCCGCCCCGC	GCCAGCCCCT	GGACGCCCGC	GACCGGCGCA	CGGATGTCGT	34020
	GATCACGGGC	ACCCGCGCCC	CCAGACCGAT	GGCCGGGACC	GGGGCCGGGG	GCGCGGGGGC	34080
	CAAGCGGGCC	ACCGTCAGCG	AGTTCTGTGCA	AGTGAAGCAC	ATCGACCGTG	TTGTGTCCCC	34140
	GAGCGTCTCT	TCCGCCCCCC	CGCCGAGCGC	CCCCGACGCG	AGTCTGCCGC	CCCCGGGGCT	34200
	CCAGGAGGCC	GCCCCGCCGG	GCCCCCGCT	CAGGGAGCTG	TGGTGGGTGT	TCTACGCCGG	34260
35	CGACCGGGCG	CTGGAGGAGC	CCCACGCCGA	GTCCGGATTG	ACGCGCGAGG	AGGTCGCGCG	34320
	CGTGTCATGG	TTCCGGGAGC	AGGCGTGGA	GCTGTTTGGG	TCGGTGGGGG	CTCCGCGGGC	34380
	GTTTCTCGGG	GCCGCGCTGG	CCCTGAGCCC	GACCCAAAAG	CTCGCCGTCT	ACTACTATCT	34440
	CATCCACCGG	GAGCGGCGCA	TGTCCCCCTT	CCCCGCGCTC	GTGCGGCTCG	TCGGTCCGGTA	34500
	CATCCAGCGC	CACGGCCTGT	ACGTTCCCGC	GCCCGACGAA	CCGACGTTGG	CCGATGCCAT	34560
40	GAACGGGCTG	TTCCGCGACG	CGCTGGCGGC	CGGGACCGTG	GCCGAGCAGC	TCCTCATGTT	34620
	CGACCTCCTC	CCGCCCAAGG	ACGTGCCGGT	GGGGAGCGAC	GCGCGGGCCG	ACAGCGCCGC	34680
	CCTGCTGCGC	TTTGTGGACT	CGCAACGCCT	GACCCCGGGG	GGGTCCGTCT	CGCCCGAGCA	34740
	CGTCATGTAC	CTCGGCGCGT	TCCTGGGCGT	GTTGTACGCC	GGCCACGGAC	GCCTGGCCGC	34800

	GGCCACGCAT	ACCGCGCGCC	TGACGGGCGT	GACGTCCCTG	GTCCTGACCG	TGGGGGACGT	34860
	CGACCGGATG	TCCGCGTTTG	ACCGCGGGCC	GGCGGGGGCG	GCTGGCCGCA	CGCGAACCGC	34920
	CGGGTACCTG	GACGCGCTGC	TTACCGTTTG	CCTGGCTCGC	GCCCAGCACG	GCCAGTCTGT	34980
	GTGAGATATC	CCAATAAAGT	GCAGTCGTTT	TCTAACCCAC	GGATGCCGTT	GTATGCCTAT	35040
5	ACGGGGGACT	ATGGGGGGGG	GGGAAAGGA	AAGGAAACAG	GAATGGAGAA	GGGAAAGGAA	35100
	CAGAGGCGGT	AGCGGACGCA	CGGCGGACAC	AATAACAAAC	AGACCGCGGA	CACGGAGGGA	35160
	GTCGGTTGGG	TTGGGCGTGG	ACGCCGCTGC	GTCCACACAC	CCGTTTATTC	GCGTCTCCAC	35220
	AAAAATGGGA	CGCACGTTCG	GACCACCCTG	AGGATGCCCG	CCAGGGCCGC	GGTGATCATA	35280
	ACGACCCCCA	GCGCGGACGC	GGCCAGAAAC	CCGGGGGCGA	TGGTGGCGAT	GGGCAGCGTG	35340
10	TCAAAGGCCA	GCAGATGAAT	CACAGTTCCG	TTGGGGAAAC	ACAACAGGGC	CACGGACGGC	35400
	ACGTCGCTGG	AAAACACGTT	CGGGGTGCCC	GCCACCGGCC	CCTGGGCCAG	CTGCTGTTGG	35460
	GTGGCATCCG	TGTCCACCAG	CAGCACCGAC	ATGACCTCCC	CGGCCGGGGT	GTAGCGCAGA	35520
	AACACGGCCC	CCACGAGGCC	GAGGTCGCGC	CGGTTTTCGG	TGCGCACACG	CCGCTTCGGC	35580
	TCAATCTCCC	GCGCGTGCCC	TTCGCGAGGT	GCGGTGAGAT	AGGTGATAAA	CAGCGGGCGG	35640
15	CGGACGTCAA	CGCCCCTAAG	CTTGTTATCCG	ATCCC CGCGG	GCAAGGGGGT	GTGGGTGACG	35700
	ACGTAGCTGG	CGTTGTGGGT	GATGGGCACG	AGGATCCGGG	GCTCCGCGTT	GTGCGACGGG	35760
	CCGCTACACT	GGTGGGTGGC	CTCCGGGACG	AAGGCGCGGA	TCAGGGCGTT	GTAGTGCGCC	35820
	CAGCGCGTGA	GAACGGAGGC	CACGCCGCGG	GTCTGTTGTG	CCATGACGTC	CGCCGGGATG	35880
	TCGGATCGGG	TGGCCATGGC	CAGCGCGTCC	AGGATGAACC	CGCCCTCGGC	GAGATCGAAG	35940
20	CGCAGGGAAG	CTGCGCATGG	GGAAAAGTGG	TCCGGGAGCC	AGAAGAGGTT	TTTCTGGTGG	36000
	TCGGTCTTGG	CTAGCGCGGC	CCGGAGATCG	GCGTGGGTG	CCGCGCGGAC	GTCGGACGTA	36060
	CACAGGGCCG	TGGTTATGAG	GAGGCCCCCG	CGGGCGCGTT	CCCCTGCTC	GGCCGAGGGC	36120
	GCGCCCCCCA	GGAACGGCGC	CCGGAGGACG	GCCGTGGCGT	AAAACAGCGC	TCGGCGGACC	36180
	ATCGGGGCGG	TTAGCGCGCG	GCCGCCGAGA	AACTCGGCGT	ACAGGGCGTC	GATCAGGCGG	36240
25	GCCGCGCTCG	GGGCCACCGC	GCCATAGGCC	GCGGGGCTGT	CCAACACGAA	CGCCAGCTGA	36300
	TAGCCAGCG	CGTGCGCCGC	CAGGCTCTGC	TCTCGCTCGA	GGATCGCGGC	CACCAGATGC	36360
	CCGAGGCGCG	CCTCCAGCCG	CAGGCGGGCC	GCCGGGTCCA	ACACGGACAC	GTTTCAGGAAC	36420
	ACCGAGTCGG	CCGCGCAGCC	CGCTGCTCCC	CGGGCGGCCA	GGCCGGCCAG	CACGCGCGAG	36480
	TGGGCCAAAA	AGCCCAGCAG	GTCGGAGAGG	CGAATCGCGT	CGTGGGCGTG	GGCCGCGTTG	36540
30	ACGAACGCAA	ACCCCAGCGA	GGCGAGCAGC	CCCGCGAGGC	GCCAGAACAG	GGACGGACGC	36600
	GCGTCCGTGC	CGGAGCCCGG	GTCTTCCCCC	AAAAACTCCG	CATAGGCCCG	CGACATATAC	36660
	TGGGCGTAGT	TCGTGCTCTC	CTCGGGGTAG	CCGGCCACCC	GCCGAGGGC	GTCCAGCGCC	36720
	GAGCCGTTGT	CGGCGGGCGT	CGGGGCCCCC	AGGACAAAGA	CGCGATACCT	GGGGCCGGCC	36780
	GGAGGCCCCG	GGAGCACC GC	GGGGGCGTTT	TCGTGCGTGC	GATTTCGGAC	CCGAGCGAGG	36840
35	GTCTTGTTCCG	CAGGCACCAC	TATGATCTCG	GCCGGAGGGC	TGTCCCGCAT	CGATATCACG	36900
	AGCCCCATGA	AGCCCTTCCC	GTATCGCGCG	CGCACGAGCG	CGGCGTCGCA	CCCGAACGCC	36960
	AGCCCGCCCG	TCGTCCAGAC	GCCCACGGGC	CACGTCGAGG	CCGACGGGGA	GAGGTACACG	37020
	TACCGACCCG	GAGTCCGTAG	CAGGCCCCTG	GCGGCCAGCC	AGGTCACGGA	TGCGTTGTGC	37080
	AGATGCGCGA	TGCTCAGGTT	CGTCGTGCGA	TGCCTCGGTG	TCCCCGCGGG	CGGCCCCGGG	37140
40	GGCGGCGCGT	TGCGTCGGCC	GTCCGGGTGC	CTCTCGGTGC	CCCCGTCGTC	TCCCCGCGGG	37200
	AACGTAAGCC	CCTCGCGGTC	CGGCGCGGCC	GCGAATGTTA	CCCAGGCCCG	GGACCGCAAC	37260
	AGCGCGGAGG	CGCCGGGGTT	GTGCGACAGT	CCCTTGAGCT	GGGTCACCTC	GGCGGGGGGA	37320
	CGGGACGTGG	GCCCCGCCTC	GGGAGCTCG	GGCAGGCTCG	CGTTCCGAGG	CCGGCCGAGC	37380

	AGATAGGTCT	TTGGGATGTA	AAGCAGCTGC	CCGGGGTCCC	GAGGAAACTC	GGCCGTGGTG	37440
	ACCAACACGA	AACAAAAGCG	CTCGGCGTAC	CACCGAAGCA	TGGGCACGGA	TGCCGTAGTC	37500
	AGGTTGAGTT	CGCCCCGGGG	CGCCAAGCGT	CCGCGCTGGG	GGTCGCTGGC	GTCGGGGGTG	37560
	TTGGGCAACC	ACAGACGCCC	GGTGTTTGTG	TCGCGCCAGT	ACGTGCGGGC	CAACCCAGCA	37620
5	CCGTGCAAAA	ACCACGGGTC	GATTTGTCTC	GTCCAGTACG	TGTCATGGCC	CCCGGCAACG	37680
	CCCACCAGGA	CCCCCATCAC	CACCCACAGA	CCGGGGCCCA	TGGTCGTCCG	TCCCCGGTGC	37740
	CAGTCCGCAG	ATGGGGGGGT	GTCCGTACCC	ACGGCCCAAA	GAGGCTCCGC	ACCTCGGAGG	37800
	CTATCGGAGG	CCCTTTGTGT	CCGTAAGCGC	GGGCCAAAAG	ATGGGGTGGG	GTGAGGGTAA	37860
	AAGACAAAAG	GGAGTACCAG	ACCGAAAACA	AGGACGGATC	GGCCCGCTCC	GTTTTTCGGT	37920
10	GGGGTGCTGA	TACGGTGCCA	GCCCTGGCCC	CGAACCCCGG	CGCTTATGGA	CACACCACAC	37980
	GACAAACAATG	CCTTTTATTC	TGTTCTTTTA	TTGCCGTGTC	CGCCGGGAGG	CCTTCCGPTC	38040
	GGGCTTCCGT	GTTTGAAC TA	AACTCCCCC	ACCTCGCGGG	CAAACGTGCG	CGCCAGGTGC	38100
	CGTATCTCGG	CGATGGACCC	GGCGGTTGTG	ACGCGGGTTG	GGATCATCCC	GGCGGTGAGG	38160
	CGCAACAGGG	CGTCTCGACA	CCCGACGGGC	GACTGATCGT	AATCCAGGAC	AAATAGATGC	38220
15	ATCGGAAGGA	GGCGGTCCGC	CAAGACGTCC	AAGACCCAGG	CAAAAATGTG	GTACAAGTCC	38280
	CCGTTGGGGG	CCAGCAGCTC	GGGAACCGCG	AACAGGGCAA	ACAGCGTGTC	CTCGATCGCG	38340
	GGCAGAGACC	CCGCGCCGTC	CTCGGGGTGC	GGGCGCGGGG	TCGCCGCGGC	GACCCCCGTC	38400
	AGCCGGCCCC	AGTCTCTCCG	CCACCTCCCG	CCGCGCTGCA	GGTACCGCAC	CGTGTGGCG	38460
	AGTAGATCGT	AGACACGGCG	AATGGCGGAC	AGCATGGCCA	GGTCAAGCCG	CTCGCCCGGG	38520
20	CGTTGGCGTC	TGGCCAGGCG	GTGCGGTGTG	TCGGCCTCCG	GAAGGACACC	CAGGACCAGG	38580
	TTCGTGCCGG	GCGCGGTCCG	GGGCATGAGG	GCCACGAACG	CCAACACGGC	CTGGGGGGTC	38640
	ATGCTTCCCA	TGAGGTACCG	CGCGGCCGGG	TAGCACAGCA	GGGAGGCGAT	AGGGTGCCGG	38700
	TCGAAAACAA	GGGTGAGGGC	CGGGGGCGGG	GCTTGCGGGC	CCACAGCCTC	CCCCCCGATA	38760
	TGAGGAGCCA	AAACGGCGTC	CGTCGCCGCA	TAAGGCGTGC	TCATTGTTAT	CTGGGCGCTG	38820
25	GTCATTACCA	CCGCCGCCTC	CCCGGCCGAT	ATCTCGCCGC	GGTCCAGACG	GTGCTGCGTG	38880
	TTGTAGATGT	TCGTCAGGGT	CTCGGAGGCC	CCCAGCACCT	GCCAGTAAGT	CATCGGCTCG	38940
	GGGACGTAGA	CGATATTGTC	GCGCGGCCCC	AGGGCCTCCA	TCAGCTGCGC	GGAGGTGGTG	39000
	GTCTTCCCCA	CCCCGTGGGG	TCCGTCTATA	TAAACCCGCA	GCAGCGTGGG	CAGCTCCGGA	39060
	TCCCCGCGGG	CTTCGGAGGC	CCCCTGCCGA	TGGCTAGGAC	GGGACGCCGC	GCGGCCGTCG	39120
30	GTAGGCCCGC	TCGCACGAGC	AGCCTGACCG	AACGCAGGCG	CGTGCTGTTG	GCCGGCGTGA	39180
	GAAGCCATAC	CCGCTTCTAC	AAGGCGTTCG	CCCGAGAGGT	GCGGGAGTTC	AACGCCACCA	39240
	GGATTTGTGG	AACGCTGCTG	ACGCTGATGA	GCGGGTCGCT	GCAGGGTCGC	TCGCTGTTTCG	39300
	AGGCCACGCG	CGTCACCTTA	ATATGCGAAG	TGGACCTCGG	GCCGCGCCGC	CCAGACTGCA	39360
	TCTGCGTGTT	CGAATTCCGC	AATGACAAAA	CGTTGGGAGG	TGTGTGCGTC	ATCCTGGAGC	39420
35	TAAAGACATG	CAAATCGATT	TCTTCCGGGG	ACACGGCCAG	CAAACGCGAA	CAGCGGACCA	39480
	CGGGCATGAA	GCAGCTGCGC	CACTCCCTGA	AGCTGCTGCA	GTCGCTCGCG	CCTCCGGGGG	39540
	ACAAGGTCGT	CTACCTGTGT	CCTATTTTGG	TGTTTGTGCG	GCAGCGTACG	CTGCGCGTCA	39600
	GCCGCGTGAC	CCGGCTCGTC	CCGCAAAAAG	TCTCCGGCAA	CATCACC GCG	GCCGTGCGGA	39660
	TGCTCCAAAG	CCTGTCCACG	TATGCCGTGC	CGCCGGAACC	GCAGACCCGG	CGGTGCGGGC	39720
40	GCCGGGTGCG	CGCGACCGCC	AGACCGCAAA	GGCCCCCTC	CCCGACACGT	GACCCGGAAG	39780
	GCACGGCGGG	TCATCCGGCC	CCACCAGAGA	GCGACCCCCC	CTCCCCAGGG	GTCTGTAGGCG	39840
	TCGCTGCGGA	GGGTGGGGGT	GTGCTTCAGA	AAATCGCGGC	GCTTTTTTGC	GTGCCGTTGG	39900
	CCGCCAAGAG	CAGACCCCGG	ACCAAAACCG	AGTGAGGTTT	TGTGTGTTGT	TTTTTTTCCT	39960

	CGTTTTGTTT	TCTCTTCTTT	CCCCCCCCC	TCCCCCGCTT	CTGGCCAAGC	ATCCTCACCT	40020
	GCTTAAGCGG	AACCCGCGGG	CGCGCGGGGA	CTCATTTGTC	GCCGGCGACA	CCCACCCGAC	40080
	AACAGCCCCCT	GGGTGTAGAC	CGCTGTCGCC	CCCGTCTGTC	GCCTCTCCCT	TTTTTCCCCC	40140
	CCTCAAAGAA	CGTGGTGTG	GGCGCCGGCC	AATTCTTCCC	GGAGCGCCGT	CGTCGCCCCG	40200
5	CCGCCGCCCT	CGAACATGGA	CCCGTACTAC	CCTTTCGACG	CGCTGGACGT	TTGGGAACAC	40260
	AGGCGCTTCA	TCGTCGCCGA	CTCCAGGAGC	TTCATCACCC	CCGAGTTCCC	CCGGGACTTC	40320
	TGGATGTTGC	CCGTGTTCAA	CATCCCCCGG	GAGACGGCGG	CGGAGCGGGC	GGCAGTGCTG	40380
	CAGGCCCAGC	GCACCGCGGC	CGCGGCGGCC	CTGGAGAACG	CCGCCCTCCA	GGCCGCGGAG	40440
	CTGCCCCTCG	ACATCGAGCG	CCGGATACGC	CCGATCGAGC	AGCAGGTGCA	TCACATCGCC	40500
10	GACGCCCTGG	AGGCGCTGGA	GACCGCGGCG	GCCGCGGGCC	AAGAGGCGGA	TGCCGCGCGG	40560
	GACGCCGAGG	CGAGGGGGGA	GGGCGCTGCG	GACGGGGCAG	CGCCGTCGCC	CACCGCGGGC	40620
	CCCGCCGCCG	CGGAGATGGA	GGTTCAGATC	GTACGCAACG	ACCCGCCGCT	ACGATACGAT	40680
	ACCAACCTCC	CCGTGGATCT	GCTACACATG	GTGTACGCGG	GCCGCGGGGC	CGCGGGTTCG	40740
	TCGGGAGTCG	TCTTTGGTAC	CTGGTACCGC	ACGATCCAGG	AACGCACCAT	CGCGGACTTC	40800
15	CCCCTGACCA	CCCGCAGCGC	CGACTTTCGA	GACGGGCGCA	TGTCCAAGAC	CTTCATGACC	40860
	GCGCTGGTCC	TGTCTCTGCA	GTCTGTGCGC	CGGCTGTACG	TGGGCCAGCG	CCACTATTCC	40920
	GCCTTCGAGT	GCGCCGTGCT	GTGTCTGTAT	CTGCTGTACC	GAACCACCCA	CGAGTCCTCC	40980
	CCCGATCGCG	ATCGCGCTCC	CGTTGCGTTC	GGGGACCTGC	TGGCCCGCCT	GCCGCGCTAC	41040
	CTGGCGCGTC	TGGCCGCGGT	AATCGGCGAC	GAGAGCGGAC	GCCCGCAGTA	CCGCTACCGC	41100
20	GACGACAAGC	TGCCCAAAGC	GCAGTTCGCG	GCGGCCGGCG	GCCGCTACGA	GCACGGGGCC	41160
	CTGGCCACCC	ACGTCGTGAT	CGCCACGTTG	GTGCGCCACG	GGGTGCTACC	GGCGGCCCCG	41220
	GGCGACGTTT	CCCGAGACAC	CAGCACCCGC	GTGAACCCCG	ACGACGTGGC	CCACCGCGAC	41280
	GACGTCAACC	GCGCCGCCGC	CGCGTTTGTG	GCACGCGGCC	ACAACCTCTT	CCTGTGGGAG	41340
	GACCAGACGC	TGCTGCGGGC	GACCGCCAAC	ACCATTACGG	CCCTGGCCGT	GCTTCGGCGG	41400
25	CTCCTCGCGA	ACGGCAACGT	GTACGCGGAC	CGCCTCGACA	ACCGCCTGCA	GCTGGGCATG	41460
	CTGATCCCGG	GAGCCGTCCC	GGCGGAGGCC	ATCGCTCGGG	GGGCGTCCGG	ATTGGACTCG	41520
	GGCGCCATAA	AAAGCGGCGA	CAACAACCTG	GAGGCGCTGT	GCGTTAACTA	TGTACTTCCG	41580
	CTGTATCAGG	CAGACCCAC	GGTCGAGCTG	ACCCAGTTGT	TTCCGGGGCT	GGCCGCCCTG	41640
	TGCCTGGACG	CCCAGGCGGG	GCGGCCACTG	GCGTCGACGA	GGCGCGTGGT	GGATATGTCG	41700
30	TCGGGCGCCC	GCCAGGCGGC	GCTCGTGCGC	CTACCGCGC	TGGAGCTCAT	CAACCGCACC	41760
	CGCACAAACA	CCACCCCTGT	GGGGGAGATT	ATTAACGCCC	ACGATGCCCT	GGGGATACAA	41820
	TACGAACAGG	GCCTGGGGCT	GCTCGCCCAG	CAGGCACGCA	TCGGCTTGGC	GTGGAACGCC	41880
	AAGCGATTTC	CCACGTTCAA	CGTGGGCAGC	GACTACGACC	TGTTGTACTT	TTTGTGTCTC	41940
	GGGTTTCATT	CCCAGTACCT	GTCCGTGGCC	TAGGGAAGGG	TGGGGGTGGT	GGTGGTGGGG	42000
35	TGTTTTTCTG	CTGTTGTTGT	TTCTGGTCCG	CCTGGTCACA	AAAGGCACGG	CGCCCCGAAA	42060
	CGCGGGCTTT	AGTCCCGGCC	CGGACGTCGG	CGGACACACA	ACAACGGCGG	GCCCCGTGGG	42120
	TGGGTAAGTT	GGTTCGGGGG	CATCGCTGTA	TTCCCTTGCC	CGCTTCCACC	CCCCCTTCCC	42180
	GTTTTGTTTG	TTTGTGCGGG	TGCCCATGGC	GTCGGCGGAA	ATGCGCGAGC	GGTTGGAGGC	42240
	GCCTCTGCCC	GACCGGGCGG	TGCCCATCTA	CGTGGCCGGG	TTTTTGGCCC	TGTACGACAG	42300
40	CGGGGACCCG	GGCGAGCTGG	CCCTGGACCC	AGACACGGTG	CGTGCGGCCC	TGCTCCGGA	42360
	GAACCCCTTG	CCGATCAACG	TAGACCACCG	CGCTCGGTGC	GAGGTGGGCC	GGGTGCTCGC	42420
	CGTGGTCAAC	GACCCCTCGG	GGCCGTTTTT	TGTGGGGCTG	ATCGCGTGCG	TGCAGCTGGA	42480
	GCGCGTCCCT	GAGACGGCCG	CCAGCGCCCG	TATTTTTGAG	CGCCGCGGAC	CCGCGCTCTC	42540

	CCGGGAGGAG	CGTCTGCTGT	ACCTGATCAC	CAACTACCTG	CCATCGGTCT	CGCTGTCCAC	42600
	AAAACGCCGG	GGGACGAGG	TTCCGCCCGA	CCGCACCCTG	TTTGCGCACG	TGGCCCTGTG	42660
	CGCCATCGGG	CGGCGCCTTG	GAACCATCGT	CACCTACGAC	ACCAGCCTAG	ACGCGGCCAT	42720
	CGCTCCGTTT	CGCCACCTGG	ACCCGGCGAC	GCGCGAGGGG	GTGCGACGCG	AGGCCGCCGA	42780
5	GGCCGAGCTC	GCGCTGGCCG	GGCGCACCTG	GGCCCCCGGC	GTGGAGGCGC	TCACACACAC	42840
	GCTGCTCTCC	ACCGCCGTCA	ACAACATGAT	GCTGCGTGAC	CGCTGGAGCC	TCGTGGCCGA	42900
	GCGGCGGCGG	CAGGCCGGGA	TCGCCGGACA	CACGTACCTT	CAGGCGAGCG	AAAAATTTAA	42960
	AATATGGGGG	GCGGAGTCTG	CCCCTGCGCC	GGAGCGCGGG	TATAAAACCG	GCGCCCCGGG	43020
	TGCCATGGAC	ACATCCCCCG	CCGCGAGCGT	TCCCGCGCCG	CAGGTGCGCG	TCCGTGCGCG	43080
10	TCAAGTCGCG	TCGTGTCGT	CTTCTTCTTC	TTCTTTTCCG	GCACCGGCCG	ATATGAACCC	43140
	CGTTTTCGCA	TCGGGCGCCC	CGGCCCTTCC	GCGCCCCGGC	GACGGGAGTT	ATTTGTGGAT	43200
	CCCCGCCTTT	CATTACAATC	AGCTCGTCAC	CGGGCAATCC	GCGCCCCACC	ACCCGCCGCT	43260
	GACCGCGTGC	GGCCTGCCGG	CCGCGGGGAC	GGTGGCCTAC	GGACACCCCG	GCGCCGGCCC	43320
	GTCCCCGCAC	TACCCGCCTC	CTCCCCGCCA	CCCCTACCCG	GGTATGETGT	TCGCGGGCCC	43380
15	CAGTCCCTTG	GAGGCCCAGA	TCGCCGCGCT	GGTGGGGGCC	ATCGCCGCCG	ACCGCCAGGC	43440
	GGGTGGGCTT	CCGGCGGCCG	CCGGAGACCA	CGGGATCCGG	GGGTGCGCGA	AGCGCCGCCG	43500
	ACACGAGGTG	GAGCAGCCGG	AGTACGACTG	CGGCCGTGAC	GAGCCGGACC	GGGACTTCCC	43560
	GTATTACCCG	GGCGAGGCCC	GCCCCGAGCC	GCGCCCGGTC	GACTCCCGGC	GCGCCGCGCG	43620
	CCAGGCTTCC	GGGCCCCACG	AAACCATCAC	GGCGCTGGTG	GGGGCGGTGA	CGTCCCTGCA	43680
20	GCAGGAACTG	GCGCACATGC	GCGCGCGTAC	CCACGCCCCC	TACGGGCCGT	ATCCGCCCGT	43740
	GGGGCCCTAC	CACCACCCCC	ACGCAGACAC	GGAGACCCCC	GCCCAACCAC	CCCGCTACCC	43800
	CGCCGAGGCC	GTCTATCTGC	CGCCGCCGCA	CATCGCCCCC	CCGGGGCCTC	CTCTATCCGG	43860
	GGCGGTCCCC	CCACCCTCGT	ATCCCCCAGT	TGCGGTACC	CCCGGTCCCG	CCCCCCCCGT	43920
	ACATCAGCCC	TCCCCCGCAC	ACGCCCACCC	CCCTCCGCGG	CCGCCGGGAC	CCACGCCTCC	43980
25	CCCCGCCGCG	AGCTTACCCC	AACCCGAGGC	GCCC GGCGCG	GAGGCCGGCG	CCTTAGTTAA	44040
	CGCCAGCAGC	GCGGCCACG	TGAACGTGGA	CACGGCCCCG	GCCGCCGATC	TGTTTGTGTC	44100
	ACAGATGATG	GGGTCCCGCT	AACTCGCCTC	CAGGATCCGG	ACTTGGGGGG	GGTGTGTGTT	44160
	TTCATATATT	TTAAATAAAC	AAACAACCGG	ACAAAAGTAT	ACCCACTTCG	TGTGCTTGTG	44220
	TTTTTGTTTG	AGAGGGGGGG	GGTGGAGTGG	GGGGGAAAGT	GGGCCGAATG	ACACAAAAAT	44280
30	TAGGTCGGAG	GGGTGAGGGG	GGGGGGGCTA	GGAGCCGAAC	CGATGGCCCC	CACACGCGAC	44340
	GGAAGGCCCG	GAAGACTACC	ACGGGGAGGG	GGTGTGGAAA	GCGACCGGTC	GCAGGGAGAC	44400
	GGGGTTGGTT	TGGGGTTGGT	TTGGGGTTGG	TTTTCCCGTT	AGCACATGTC	TGCATTTGTT	44460
	TTTCTAGTCA	CACGCCCCCC	CCCCCAAATA	AAAACCAAGG	CAAAACAATA	CCAGAAGTCA	44520
	TGTGTATTTT	TGAACATCGG	TGTCTTTTTA	TTTATACACA	AGCCCAGCTC	CCCTCCCTC	44580
35	CCTTAGAGCT	CGTCTTCGTC	TCCGGCCTCG	TCCTCGTTGT	GGAGCGGAGA	GTACCTGGCT	44640
	TTGTTGCGCT	TGCGCAGAAC	CATGTTGGTG	ACCTTGGAGC	TGAGCAGGGC	GCTCGTGCCC	44700
	TTCTTTCTGG	CCTTGTGTTT	CGTGCCTCC	ATGGCCGACA	CCAAAGCCAT	ATATCGGATC	44760
	ATTTCTCGGG	CCTCGGCCAA	CTTGGCCTCG	TCAAACCCGC	CCCCCTCCGC	GCCTTCCTCC	44820
	CCCTCCCCGC	CCACGCCCCC	GGGGTCGGAA	GTCTTGAGTT	CCTTGGTGGT	GAGCGGATAC	44880
40	AGGGCCTTCA	TGGGATTGCG	TGCAAGTTGC	AGGACGTAGC	GGAAGGCGAA	GAAGGCCGCG	44940
	ACCAGGCCGG	CCAGGACCAG	CAGCCCCACG	GCAAGCGCCC	CGAAGGGGTT	GGACATAAAG	45000
	GAGGACACGC	CCGAGACGGC	CGACACCACG	CCCCCACTA	CTCCCATGAC	TACCTTGCCG	45060
	ACCGCGCGCC	CCAAGTCCCC	CATCCCCTCG	AAGAACGCGC	ACAGCCCCGC	GAACATGGCG	45120

	GCGTTGGCGT	CGGCGCGGAT	GACCGTGTCTG	ATGTCGGCAA	AGCGCAGGTC	GTGCAGCTGG	45180
	TTGCGGCGCT	GGACCTCCGT	GTAGTCCAGC	AGGCCGCTGT	CCTTGATCTC	GTGGCGCGTG	45240
	TAGACCTCCA	GGGGCACAAA	CTCGTGGTCC	TCCAGCATGG	TGATGTTTCA	GTCGATGAAG	45300
	GTGCTGACGG	TGGTGACGTC	GGCGCGACTC	AGCTGGTGAG	AGTACGCGTA	CTCCTCGAAG	45360
5	TACACGTAGC	CCCCGCCGAA	GATGAAGTAG	CGCCGGTGGC	CCACGGTGCA	CGGCTCGAGC	45420
	GCGTCGCGGG	TGAGGCGCAG	CTCGTTGTTC	TCGCCCAGCT	GCCCCTCGAT	CAGCGGGCCC	45480
	TGGTCTTCGT	ACCGAAAGCT	GACCAGGGGG	CGGCTGTAGC	ACGTCCCCGG	CCGCGAGCTG	45540
	ACGCGCATCG	AGTTCTGCAC	GATCACGTTG	TCCGGGGCGA	CGGGCACGCA	CGTGGAGACG	45600
	GCCATGACGT	CTCCGAGCAT	GCGCGCGCTC	ACCCGCCGGC	CGACGGTGGC	GGAGGCGATG	45660
10	GCGTTGGGGT	TGAGCTTGCG	GGCCTCGTTC	CAGAGAGTCA	GCTCGTGGTT	CTGCAGCTCG	45720
	CACCACGCGA	CGGCGATGCG	CCCCAGCATG	TCGTTTACGT	GGCGCTGTAT	GTGGTTATAC	45780
	GTAAACTGCA	GCCGGGCGAA	CTCGATCGAG	GAGGTGGTCT	TGATGCGCTC	CACGGACGCG	45840
	TTGGCGCTGG	GCGCCTCCCG	CAGTGGCGCG	GGCGTGGCAT	TCCGGGGCTT	GCGGTCCTGC	45900
	TCCCGCATGT	ACTCCCGCAC	GTACAGCTCG	GCGAGCGTGT	TGCTGAGGAG	GGGCTGGTAC	45960
15	GCGATGAGGA	AGCCCCCGT	GGCCAGGTAG	TACTGCGGCT	GGCCCCACCT	GATGTGCGTG	46020
	GCGTTGTA	TGCGCGCAA	CATGCGGTCG	ATGGCCTCGC	GGGCATCCCG	GCCAATGCAG	46080
	TCGCCCAGGT	CGACGCGCGA	GAGCGAGTAC	TGGGTCAGGT	TGGTGGTGAA	GGTGGTCGAG	46140
	ATGGCGTCGG	AGGAGAAGCG	GAAGGAGCCG	CCGTACTCGG	CGCGGAGCAT	CTCGTCCACC	46200
	TCCTGCCACT	TGGTCATGGT	GCAGACCGCC	GGTCGCTTCG	GCACCCAGTC	CCAGGCCACG	46260
20	GTAAACTTGG	GGGTCGTGAG	CAAGTTGCGG	GTCGTCGGCG	ACGTGGCCCG	GGCCTTCGTG	46320
	GTGAGGTGCG	GCGCGTAGAA	GCCGTCGACC	TGCTTGAAGC	GGTCGGCGGC	GTAGCTGGTG	46380
	TGCTCGGTGT	GCGACCCCTC	CCGGTAGCCG	TAAAACGGGG	ACATGTACAC	AAAGTCGCCC	46440
	GTCGCCAACA	CAAACATC	GTACGGGTAC	ACCGACGCG	CGTCCACCTC	CTCGACGATG	46500
	CAGTTGACCG	TCGTGCCGTA	CCGATGGAAC	GCCTCCACCC	GCGAGGGGTT	GTACTTGAGG	46560
25	TCGGTGGTGT	GCCACCCCG	GCTCGTGCGC	GTGGCGACCT	TCGCCGGCTT	GAGCTCCATG	46620
	TCGGTCTCGT	GGTCGTCCCG	GTGAAACGCG	GTGGTCTCCA	TGTTGTTC	CACGTACTTG	46680
	GCCGTGGAGC	GGCAGACCCC	CTTGGCGTTA	ATCTTGTCGA	TCACCTCCTC	GAAGGGAACG	46740
	GGGGCGCGGT	CCTCGAATAT	CCCCATAAAC	TGGGAGTAGC	GGTGGCCGAA	CCACACCTGC	46800
	GACACGGTCA	CGTCTTTGTA	GTACATGGTG	GCCTTGAATT	TGTACGGGGC	GATGTTCTCC	46860
30	TTGAAGACCA	CCGCGATGCC	CTCCGTGTAG	TTCTGCCCCC	CCGGGCGCGT	CGGGCAGCGG	46920
	CGCGGCTGCT	CAAACATGCAC	CACCGTGGCG	CCCGTCGGGG	GCGGGCACAC	GTAAAACTGG	46980
	GCATCGGCGT	TCTCGACCTT	GATTTCCCGC	AGGTGCGCGC	GCAGCGTGGC	GTGGCCGGCG	47040
	GCGACGGTCG	CGTTGGCGTC	GGGGGGCGGG	GTCGCCTCGG	GCCGCTTGGG	CGGCTTTT	47100
	GTTTTCGCT	TCCGGGCCCT	GGTGGTCGCG	GGGCTCGGGA	CGGGGGGCGG	CCGGGAGGCG	47160
35	GGACCCCGT	TCGCCGCGAC	GGTCGCGGCC	ACGCCGCCCG	AGGCGCGGGG	GGCCGCCGGG	47220
	GCCGCCGGGG	CCGCCGACGC	CACCGCGGCC	ACCAGCGCCC	CCACGACCAG	CGCGCAAATC	47280
	AAGCCCCCCC	CGCGCATGGC	GGGCTTACGG	GGGCGCGTCG	CTCCCGCCGC	CCGCTAGTCT	47340
	GGGGGCGAGG	TGCTGCAGGA	CCGAGTAGAG	GATGGAAAAA	ACGTCTCGGT	CGTAAACCAC	47400
	GACCGAGCGG	GGTCCGATGC	AGCCGTCGGG	GCCGCTCTCG	ACGATGGCCA	CCAGCGGACA	47460
40	GTCCGAGTTG	TACGTGAGGT	ACACGCCCGG	CGGGTAGCGG	TACAGACCTT	CGGAGGTGCG	47520
	GCGGCTGCAG	TCGGGGCGGC	GCAACTCAAG	CTCCCCGCAC	CGGTAGACCG	ACGCAAAGAG	47580
	TGTGGTGGCG	ATAATGAGCT	CGCGAATATA	TCGCCAGGCG	GCGCGCTGGG	TGGGCGTGAT	47640
	TCCGGAACA	CCGTCAAAAC	AGTAGAACTT	TTGAAACTCG	CTGACGGCCC	AATCAGCGCC	47700

	CGAACCCCC	GCGCCCATGA	TGAAGCGGGC	GAGTTCTCTC	TTGAGGTGCG	GCAGGAGCCC	47760
	CACGTTCCTCG	ACGCTGTAGT	ACAGCGCGGT	GTTGGGGGGC	TGGGCGAAGC	TGTGGGTGGA	47820
	GTGGTCGAAC	AGGGGCCCGT	TGACGAGCTC	GAAGAAGCGA	TGGGTGATGC	TGGGGAGCAG	47880
	GGCCGGGTCC	ACCTGGTGGC	GCAGCAGCGA	CGCTCGCATG	AACCGGTGCG	CGTCAAACAC	47940
5	GCCCCGGGCG	GCGCGGTGTG	CGATGACCGT	GCCCCGCGCC	GCCGTCAGGG	CGCAGAAGCG	48000
	CGCGCGCGCC	GCGAAGCCGT	TGGCGACCGC	GGCGAAGGTC	GCGGGCAGCA	CCTCGCCGTG	48060
	GACGCTGACC	CGCAGCATCT	TCTCGAGCTC	CCCGCGCTGC	TCGCGCACGC	AGCGCCCGAG	48120
	GCTGGCCAGC	GACCGCTTGG	TCAGGCGGTC	CGCGTACAGC	CGCCGGCGCT	CCGGCACGTC	48180
	CGCGGCGGCC	CGCGTCGCGA	TGTCGCCCCA	GCTCTCCGGC	CCCTGCGCCC	CTGGCTCGGG	48240
10	GCCGCGCTCC	CCGTCCTCGC	TCGCGGGCGT	CCCCGCGCCA	CGCCTCCGCC	CCCCCTCCTC	48300
	CGCGGCGGCC	CGGGGCTCTT	CCTCCTCGGC	CCCCCGGTC	GCGCCGCCGG	CCCCCAGCCG	48360
	CGCCAGCACG	CGGCGCAGCG	CCTCCTCGTC	GCCTGCTCG	GGGCTGACGA	GCCGCCGCAG	48420
	CAGCGGCGTC	GTGAGGTGGT	GGTCGTAGCA	CGCGCGTATC	AGCGCCTCGA	TCTGATCGTC	48480
	GGGCGACGTC	GCCTGGCCGC	CGATGATCAG	GGCGTCCACC	ATGTCCAGCG	CCGCCAGGTG	48540
15	GCCCCCGAAC	GCGCGATCGA	AGTGCTCCGC	CCGCCGCCCG	AACAGCGCCA	GCTCCACGGC	48600
	CACCGCGGCG	GTCTCCTGCT	GCAGCTCGCG	CTGCGCCAGC	GCGTTCAGGT	TGTCGGCGAA	48660
	GGCGTCCATG	GTGGAGTGGC	GGGCGCGATC	GCCGGACGCC	AGCCAGAAGC	GCAGCTCGCT	48720
	GATGGCGTAC	AGGCCGGGCG	TAGTGGCCTG	AAACACGTCA	TGCGCCTCCA	GCAGGGCGTC	48780
	GGCCTCCTCG	CGGACAGAAG	AGCTATCGGC	GGGCGGCGGG	CCGGCCCCGG	CCCCGCCGCC	48840
20	CGCCGCGGTC	CGCGCCAGCG	CCTGGTCCAG	CACACAGAGC	GCTCGCGCGC	GGGCGGCGTC	48900
	CGACAGCCCC	GCGGCGTGGG	GCAGGTACCG	TCGCAGCTCG	TTGGCGTCCA	GCCGCACCTG	48960
	GGCCTGTTGG	GTGACGTGGT	TACAGATGCG	GTCCGCCAGG	CGGCGGGCGA	TGGTCGCCCC	49020
	CTGGTTCGCG	GTGACGCACA	GCTCCTCGAA	ACAGACCGCG	CACGGGTGGG	ACGGGTGCGT	49080
	CAGCTCCGGG	GGCACGATGA	GGCCCCAGCC	CACCGCCGCC	ACCATAAACT	CCCGGACGCG	49140
25	CTCCAGCGCG	GCCGTGGCGC	CGCTCGGGGG	GGTGATGAGG	TGGCAGTAGT	TCAGCTGCTT	49200
	GAGAAATTC	TCGACATCAT	GCAGGAAGCA	CAGTCCATG	CGGACGTCCC	CGCCGTACGT	49260
	CTCCAGCCGG	ATCTGCTGGT	GGTACGGACA	GGGTGCGGCC	AGACCCATGG	TCTCGGTGAA	49320
	AAAGGCAGAG	ACGTCACCCG	TGGTCGCGAA	CGTTTCCAGG	TGGCCCAGGA	GCCGCTCCCC	49380
	CTCGCGCCAC	GCGTACTCCA	GGAGCAACTC	CAGGGTGACC	GACAGCGGGG	TGAGAAAGGC	49440
30	GGCGGCCTGA	GCCTCCAGCC	CCGGCCGCGAG	GTGCCGCCGC	AGCACGCGCA	CCTGGAGCGC	49500
	GTTGAGCTTT	AGCTGGGCGA	GCTTCCCCAG	GCCGATCTGG	GGGTGCGATC	GTCGAAGCAG	49560
	CTCTAGCTGA	AAAACGTACG	TCTGTACCTG	CCCAGCAGG	GCCAACAGTT	TCTGTGCGGC	49620
	CGCAGTGGGC	TCGGAACCG	CGGCCGGGGG	CGCGGCCGCC	ATGGCGAGTC	GCCCCGCCGT	49680
	GCTGTGGTTT	AGTTAAGGTT	TGGGGGGGTG	GGTCAGAGGC	GCGCCCCGCG	CGGACTGATG	49740
35	CGGCGGCGGG	CCCCTGACAT	CCCCTCTTTA	TGCCGTCGC	CCGCCCGCCC	GCCCCGCCGG	49800
	TGTGCCGTGA	TTGCGGAGT	CGGGGCCTTG	TGTTTCTTTC	TTTCCCCCCC	CGAATCCGTT	49860
	CTTCTCTCCT	CACCCCCCCT	CCCCACACAC	CCACCCAGGA	CTCGCCACCA	CAAGGAGGCG	49920
	AGAGCCCGTC	GCTAACCCAA	AGACACAGTC	ACGAGACAGC	ATATCGACTG	TAGTTGCGAT	49980
	CGTTTATTTT	ATACACAACA	CCAACCTTTC	CTTCGACCCC	CCCCACCCCC	GCCCCCTAGAG	50040
40	CATATCCAAC	GTGAGGTCTT	TTTTCTCCGG	TGGTCCCTCC	CCAAACGGAT	CGTCGCCGTG	50100
	AAACGCCCGC	TTTCGGGCGA	CGCCGGCCGC	CCCCGCCGCC	GCCGCCAAAC	CGCCGAACGA	50160
	CGCCGCGTGG	TCATCCTCGT	CGCCGAAATC	CCCAAAGTTA	AACACCTCCC	CGGCGGCGCC	50220
	GAGCTGGCTG	ACCAGGGCCT	CCGCCTCGTG	GGCCACCTCC	AGGGCCGCGT	CGGTGACCA	50280



	CTCGCCGTGC	CCGCGCTCCA	GGGCGCGGGT	GGTAAACTCC	ATCATTTCCCT	CGCTCAGGTA	50340
	CTCGTCCTCC	AGCAGCGCCA	GCCAGTCCTC	GATCTGCAGC	TGCTGGGTGC	GGGGGCCCAG	50400
	GCTCTTGACG	GTCGCCACAA	ACACGCTGCT	GGCGACCGCC	GCCCCGCCCT	CCGCAATGAT	50460
	GCCCCGAGC	TGCTCGCACA	GCGAATGCTC	GTGGGCCCCG	CCCCCGAGAC	TCGACGCCGC	50520
5	GCACACAAAC	CCGGCCCTGG	GGCAGGCCAG	GACAAACTTG	CGGGTGCGGT	CAAAGATCAG	50580
	CAGCGGGCAC	GCGTTTTTGC	CGCCCAGCAG	GCTGGCCCAG	TTCCCCGGCCT	GAAACACGCG	50640
	GTCGTTGCCG	GCCATGCCGT	AGTATTTGCT	GATGCTGAGG	CCCAGCACGA	CCATCGGGCG	50700
	CGCGGCCATC	ACGGGCCGCA	GCAGGTTGCA	GCTCGCGAAC	ATGGACGTCC	AGGCGCCGGG	50760
	GTGCGCGTCG	AGGGAGTCCA	TCAGCGCGCG	GGCCCCGGCC	TCCAGGCCCG	CGCCGCCCTG	50820
10	CGGGGCCCAG	GCGGCGGCCG	CCTGCACGCC	GGGGGACGG	CGGGACCCGG	CGATGACGGC	50880
	CGTGAGGGTG	TTTATGAAGT	ACGTCGAGTG	GTCGCAGTAC	CTCAAGATCT	GGTTGGCCAT	50940
	GTAGTACATG	GCCAGTTCGC	TCACGTTATT	GGGGGCCAGG	TTGATAAAGT	TAATCGCGCC	51000
	GTAGTCCAGG	GAGAACCTCT	TAATGAACGC	GATGGTCTCT	ATGTCTTCGC	GCGACAAGAG	51060
	CCGGGCGGGG	AGCTGGTTGC	GCTGGAGGGC	GGTCCAGAAC	CACTGCGGGT	TCGGCTGGTT	51120
15	CGACCCCGGG	GGCTTGCCGT	TGGGAAAGAT	GACCGCGTGG	AACTGCTTCA	GCAGGAAGCC	51180
	CAGCGGTCCG	AGGAGGATGT	CCACGCGCTT	GTCGGGCTTC	TGGTAGGCGC	TC'TGGAGGCT	51240
	GGCGACCCGC	GCCTTGCGCG	CCTCGGACGC	GTTGGCGCTC	GCGCCCGCGA	ACAACACGCG	51300
	GCTCTTGACG	CGCAGCTCCT	TGGGAAACCC	CAGGGTCACG	CGGGCAACGT	CGCCCTCGAA	51360
	GCTGCTCTCG	GCGGGGGCCG	TCTGGCCGGC	CGTTAGGCTG	GGGGCGCAGA	TAGCCGCCCC	51420
20	CTCCGAGAGC	GCGACCGTCA	GCGTCTTCGC	CGACAGGAAC	CCGTGTGTTGA	ACAGGTCCAT	51480
	GACGCGCCGC	CGCAGCACCG	GTGGAATTG	ATTGCGAAAG	TTGCGCCCCCT	CGACCGACTG	51540
	CCCGGCGAAC	ACCCCGTGCC	ACTGGCTCAG	GGCCAGGTCC	TGGTACACGG	CGAGGTTGGA	51600
	CCGCCGCGCG	AGGAGCTGCA	GCAGGGGGCA	CGGCCCGCAG	GTGTACGGGT	CCAGCGACAG	51660
	CGACATGGCG	TGGTTGGCCT	CGGCCAGACC	GTCGCGGAAC	TTAAACCAGC	TGCTTGATGT	51720
25	TGTTCAACAC	CGTGTGCAGG	GCCTCGCGGG	TGCCGATAAT	CGTCTCCAGC	CTCCCCAGGG	51780
	CCGTGGGCAC	CGCCTGGTCC	ACGTACTGCA	GGGCCTCGAG	CTCGGCCATG	ACGCGCTCGG	51840
	TGGCCGCGCG	GTACGTCTCC	TGCATGATGG	TCCGGGTGTT	CTCGGACCCG	TCCGCGCGCT	51900
	TCAGGGCCGA	GAAGGCGGCG	TAGTTCCCCA	GCACGTCGCA	GTCGCTGTAC	GCGCTGTTCA	51960
	TCGTTCCGAA	GACCCCAATG	GCCCCCGGG	CGGCGCTCGC	GAAC'TGGGG	TGGCGGGCCC	52020
30	GCAGCCGCAT	CAGCGTCGTG	TGCGCGCAGG	CGTGGCGGGT	CTCGAAGGTA	CACAGGTTGC	52080
	AGGGCACGTC	GGTCTGGCCC	GAGTCCGCGA	CGTAGCGAAA	CACGTCCATC	TCCTGGCGCC	52140
	CGACGATGAC	TCCGCCGTCG	CAGCGCTCCA	GGTAAACAG	CATCTTGGCC	AGCAGGGCCG	52200
	GAGAGAACCC	GCACAGCATG	GCCAGGTGCT	CGCCGGCGAA	CTCCTGGGTT	CCGCCGACGA	52260
	GGGGCGCCGT	GGGGCGCCCC	TCGTACCCGG	GCACCACGTG	GCCCTCGCGG	TCCAGCTGCG	52320
35	GGTTGGCCGC	CACGTGCGTG	CCGGGCACGA	GAAAGAAGCG	GTAAAAGGAG	GGCTTGCTGT	52380
	GGTCCTTGGG	GTCCGCCGGC	CCGGCGTCGT	CCACCTCGGT	CAGGTGGAGG	GCCGAGTTGG	52440
	TGCTGAACAC	CATGGCGCCC	ACGAGGCCCG	CGGCGCGCGC	CAGGTACGCC	CCGACGGCGC	52500
	CGGCACGGGC	CGCGGGCGTT	TCCTGGCCCT	CAAGCAGGGG	CCACGTGGTG	ATGTCGGGGG	52560
	GCGGCTCGTC	AAAGACCGCC	ATCGACACGA	TGGACTCCAG	GGCCAGGGCG	GCGTCGCCCG	52620
40	CCATCACCGA	GGCCAGGCGC	TGCTCAAACC	CGCCCCCGCG	GCCCTTGTTT	CCGGCGTCGC	52680
	GCGCGCCCCG	CTGGGGCTTA	CCCTGGCTGG	CCTCGAAGGC	CGTGAACGTA	ATGTCGGCGG	52740
	GGAGGGCCGC	GCCCTCGTGG	TTTTCGTCTGA	ACGCCAGGTG	GGCGGCCGCG	CGGGCCACGG	52800
	CGTCCACGTT	CCGAGCACGC	AGGGCCACGG	CGGCGGGCCC	GACGACCGCC	TCGAACAGCA	52860

	GGCGGGCGAG	GGGGCGGTG	AAAAACGAA	GGGGGTAGTT	GAAATTCTCC	CCGATCGATC	52920
	GGTGGTTGCA	GTAAACGGA	TCGGCGATGA	CCCGGCTAAA	ATCCGGCATA	AACATCTGCA	52980
	GCGGATACAC	GGGGATGCGG	TGAACCTCCG	CGTCCCCGAT	GGTTACCTTG	TCCATCCCGC	53040
	CCAGGTGCAG	GAAGGTGTTG	CTGATGCACA	CGGCCTCCCG	GAAGCCCTCC	GTGATCACCA	53100
5	GATACAGCAA	GGCCCGGTCC	GGGTCCAGTC	CGAGCCGCTC	GCACAGCGCG	TCCCCCGTCG	53160
	TCTCGTGCTT	TAGGTGCGAG	GGCCGGGGCG	CGTAGTCCGA	GAAGCCAAAA	TGGCGGCGCG	53220
	CCCGCTCGCA	GAGCCGCGTC	AGGTTGGGGG	CTGGGTGCT	GGGGGCCAGG	TGGCGGCGCG	53280
	CGTGAAAGAC	GTAGACGGAC	GGGCTGTAGT	GCGAGGGCAT	AAGCTTGAGG	GACACCGCGG	53340
	TCCCCCAAG	GCCCGTCGTG	CGGGACCCGA	CGACCGCGGC	CACGTTGGCC	TCAAACCCGC	53400
10	TCTCCACGGT	CAGGCCGACG	ATGAGGGGCG	CGACGGCGAC	GTCCGCGTCG	CCGCTGCGCG	53460
	CCGACAGTAG	CGACAGCAGC	TCCAGGCCTT	CGGCCGGACA	GGCGCGGCCA	TACACGTACC	53520
	CCATCGGCCC	CGGAGGAACC	TTGACGGTGG	TCGTCGTTTT	GGGCTTGGTG	TCCATGGCTT	53580
	TCGGGAGATT	GGCGACCGGC	AGGAACGGGG	GCCCGCAAG	ACGACCGGGG	GCAGACGGGG	53640
	GAGGCCGCGC	GTGGTCGACG	GCTGCTGCCC	GCCGTCGTCT	CTCCGATGGG	GTCGAATGCC	53700
15	GGCGCTGGGG	GTGGGTCTA	CACCCGCCCG	TTCACCGAGC	GGCCCTGGT	GGGGGTGGGA	53760
	TGGGTGGGAT	GGGGTGGGCG	AGAATGGCCC	GCCACCGGAT	CGCGCCGGAC	GGGGGGGCCC	53820
	GGGGTTGGGC	AAGGTTTGGG	CGCAAGGCTC	CAGCGCGCAT	TCGAGAGGCC	TGCGGATGGC	53880
	GGCCCAGAGC	TGGGTATGCT	CGGCCGGGGC	GGCCGGTATA	TGTACGGCGT	GCTGGGAGGG	53940
	GCGGCGTCGG	GCCCCGCCCA	CGGTCCGCCA	CGCCCCGCGC	GTCATCGGCA	GGGGGCGTGG	54000
20	CCGCCCTTCT	AAAAAAAGTG	AGAACGCGAA	GCGTTCGCAC	TTTGTCCTAA	TAGTATATAT	54060
	ATTATTAGGA	CAAAGTGC GA	ACGCTTCGCG	TTCTCACTTT	TTTTAGAAGG	GCGGCCACGC	54120
	CCCCTTTGAC	GTCACGCTCA	CCCGGGCGGC	CGGCCGCCCA	TAAGCGCGGC	CTGCCGGGCC	54180
	GATAAAAAGA	AACCGCGGCG	CCCCCGCGGA	CACCACACAC	TGGCTCTCGA	ACCCCGGACG	54240
	CGCAGAAGGG	ACCCGGGCGC	GGGTCCGCCG	GTAAGAGCCG	GGGGGAACAT	CGGCACCGCC	54300
25	ATCCCACCCC	GAGCTGTTGG	GTGGGCGGGT	GGGGGGGCTG	GTGAGGCGGT	GGTGGGAGGG	54360
	GGCGGCGTAT	AGCAGGACAA	CGACCGGCGG	CGATGTTTTG	TGCCGCGGGC	GGCCCGACTT	54420
	CCCCCGGGGG	GAAGTCGGCG	GCTCGGGCGG	CGTCTGGGTT	TTTTGCCCCC	CACAACCCCC	54480
	GGGGAGCCAC	CCAGACGGCA	CCGCCGCCTT	GCCGCCGGCA	GAACCTCTAC	AACCCCCACC	54540
	TCGCTCAGAC	CGGAACGCAG	CCAAAGGCCC	CCGGGCCGGC	TCAGCGCCAT	ACGTACTACA	54600
30	GCGAGTGCGA	CGAATTTCTGA	TTTATCGCCC	CGCGTTCGCT	GGAGGAGGAC	GCCCCCGCGG	54660
	AGCAGCGCAC	CGGGGTCCAC	GACGGCCGCC	TCCGGCGCGC	CCCTAAGGTG	TACTGCGGGG	54720
	GGGACGAGCG	CGACGTCCTC	CGCGTGGGCC	CGGAGGGCTT	CTGGCCGCGT	CGCTTGCGCC	54780
	TGTGGGGCGG	TGCGGACCAT	GCCCCCGAGG	GGTTCGACCC	CACCGTCACC	GTCTTCCACG	54840
	TGTACGACAT	CCTGGAGCAC	GTGGAACACG	CGTACAGCAT	GCGCGCCGCC	CAGCTCCACG	54900
35	AGCGATTTAT	GGACGCCATC	ACGCCCGCCG	GGACCGTCAT	CACGCTTCTG	GGTCTGACCC	54960
	CCGAAGGCCA	TCGCGTCGCC	GTTACGCTCT	ACGGCACGCG	GCAGTACTTT	TACATGAACA	55020
	AGGCAGAGGT	GGATCGGCAC	CTGCAGTGCC	GTGCCCCGCG	CGATCTCTGC	GAGCGCCTGG	55080
	CGGCGGCCCT	GCGCGAGTCG	CCGGGGGCGT	CGTTCGCGCG	CATCTCCGCG	GACCACTTCG	55140
	AGGCGGAGGT	GGTGGAGCGC	GCCGACGTGT	ACTATTACGA	AACGCGCCCC	ACCCTGTACT	55200
40	ACCGCGTCTT	CGTGCGAAGC	GGGCGCGCGC	TGGCCTACCT	GTGCGACAAC	TTTTGCCCCG	55260
	CGATCAGGAA	GTACGAGGGG	GGCGTCGACG	CCACCACCCG	GTTTATCCTG	GACAACCCGG	55320
	GGTTTGTAC	CTTCGGCTGG	TACCGCCTCA	AGCCCGGCCG	CGGGAACGCG	CCGGCCCAAC	55380
	CGCGCCCCCC	GACGGCGTTC	GGAACCTCGA	GCGACGTCGA	GTTTAACTGC	ACGGCGGACA	55440

	ACCTGGCCGT	CGAGGGGGCC	ATGTGTGACC	TGCCGGCCTA	CAAGCTCATG	TGCTTCGATA	55500
	TCCAATGCAA	GGCCGGGGGG	GAGGACGAGC	TGGCCTTTCC	GGTCGCGGAA	CGCCCGGAAG	55560
	ACCTCGTCAT	CCAGATCTCC	TGTCTGCTCT	ACGACCTGTC	CACCACCGCC	CTCGAGCACA	55620
	TCCTCCTGTT	TTCGCTCGGA	TCCTGCGACC	TCCCCGAGTC	CCACCTCAGC	GATCTCGCCT	55680
5	CCAGGGGCCT	GCCGGCCCCC	GTCGTCTGG	AGTTTGACAG	CGAATTCGAG	ATGCTGCTGG	55740
	CCTTCATGAC	CTTCGTCAAG	CAGTACGGCC	CCGAGTTCTG	GACCGGGTAC	AACATCATCA	55800
	ACTTCGACTG	GCCCTTCGTC	CTGACCAAGC	TGACGGAGAT	CTACAAGGTC	CCGCTCGACG	55860
	GGTACGGGCG	CATGAACGGC	CGGGGTGTGT	TCCGCGTGTG	GGACATCGGC	CAGAGCCACT	55920
	TTCAGAAGCG	CAGCAAGATC	AAGGTGAACG	GGATGGTGAA	CATCGACATG	TACGGCATCA	55980
10	TCACCGACAA	GGTCAAATC	TCCAGCTACA	AGCTGAACGC	CGTCGCCGAG	GCCGCTTGA	56040
	AGGACAAGAA	GAAGGATCTG	AGCTACCGCG	ACATCCCCGC	CTACTACGCC	TCCGGGCCCCG	56100
	CGCAGCGCGG	GGTGATCGGC	GAGTATTGTG	TGCAGGACTC	GCTGCTGGTC	GGGCAGCTGT	56160
	TCTTCAAGTT	TCTGCCGCAC	CTGGAGCTTT	CCGCCGTCGC	GCGCCTGGCG	GGCATCAACA	56220
	TCACCCGCAC	CATCTACGAC	GGCCAGCAGA	TCCGCGTCTT	CACGTGCCTC	CTGCGCCTTG	56280
15	CGGGCCAGAA	GGGCTTCATC	CTGCCGGACA	CCCAGGGGCG	GTTTCGGGGC	CTCGACAAGG	56340
	AGGCGCCCAA	GCGCCCGGCC	GTGCCCTCGG	GGGAAGGGGA	GCGCCCGGGG	GACGGGAACG	56400
	GGGACGAGGA	TAAGGACGAC	GACGAGGACG	GGGACGAGGA	CGGGGACGAG	CGCGAGGAGG	56460
	TCGCGCGCGA	GACCGGGGGC	CGGCACGTTG	GGTACCAGGG	GGCCC GG GTC	CTCGACCCCA	56520
	CCTCCGGGTT	TCACGTCGAC	CCCGTGGTGG	TGTTTGACTT	TGCCAGCCTG	TACCCAGCA	56580
20	TCATCCAGGC	CCACAACCTG	TGCTTCAGTA	CGCTCTCCCT	GCGGCCCGAG	GCCGTCGCGC	56640
	ACCTGGAGGC	GGACCGGGAC	TACCTGGAGA	TCGAGGTGGG	GGGCCGACGG	CTGTCTCTTCG	56700
	TGAAGGCCCA	CGTACGCGAG	AGCCTGCTGA	GCATCCTGCT	GCGCGACTGG	CTGGCCATGC	56760
	GAAAGCAGAT	CCGCTCGCGG	ATCCCCCAGA	GCACCCCGGA	GGAGGCCGTC	CTCCTCGACA	56820
	AGCAACAGGC	CGCCATCAAG	GTGGTGTGCA	ACTCGGTGTA	CGGGTTCACC	GGGGTGCAGC	56880
25	ACGGTCTTCT	GCCCTGCCTG	CACGTGGCCG	CCACCGTGAC	GACCATCGGC	CGCGAGATGC	56940
	TCCTCGCGAC	GCGCGCGTAC	GTGCACGCGC	GCTGGGCGGA	GTTTCGATCAG	CTGCTGGCCG	57000
	ACTTTCGGA	GCGGGCCGGC	ATGCGCGCCC	CCGGTCCGTA	CTCCATGCGC	ATCATCTACG	57060
	GGGACACGGA	CTCCATTTTC	GTTTTGTGCC	GCGGCCTCAC	GGCCGCGGGC	CTGGTGGCCA	57120
	TGGGCGACAA	GATGGCGAGC	CACATCTCGC	GCGCGCTGTT	CCTCCCCCGG	ATCAAGCTCG	57180
30	AGTGCGAAAA	AACGTTTACC	AAGCTGCTGC	TCATCGCCAA	GAAAAAGTAC	ATCGGCGTCA	57240
	TCTGCGGGGG	CAAGATGCTC	ATCAAGGGCG	TGGATCTGGT	GCGCAAAAAC	AACTGCGCGT	57300
	TTATCAACCG	CACCTCCAGG	GCCCTGGTCG	ACCTGCTGTT	TTACGACGAT	ACCGTATCCG	57360
	GAGCGGCCCG	CGCGTTAGCC	GAGCGCCCCG	CAGAGGAGTG	GCTGGCGCGA	CCCCTGCCCCG	57420
	AGGGACTGCA	GGCGTTCGGG	GCCGTCTCTG	TAGACGCCCA	TCGGCGCATC	ACCGACCCGG	57480
35	AGAGGGACAT	CCAGGACTTT	GTCCTCACCG	CCGAAGTGA	CAGACACCGG	CGCGCGTACA	57540
	CCAACAAGCG	CCTGGCCCCAC	CTGACGGTGT	ATTACAAGCT	CATGGCCCCG	CGCGCGCAGG	57600
	TCCCGTCCAT	CAAGGACCGG	ATCCCGTACG	TGATCGTGCG	CCAGACCCGC	GAGGTAGAGG	57660
	AGACGGTCCG	GCGGCTGGCC	GCCCTCCGCG	AGCTAGACGC	CGCCGCCCCA	GGGGACGAGC	57720
	CCGCCCCCCC	AGCGGCCCTG	CCCTCCCCGG	CCAAGCGCCC	CCGGGAGACG	CCGTCCGATG	57780
40	CCGACCCCCC	GGGAGGCGCG	TCCAAGCCCC	GCAAGCTGCT	GGTGTCCGAG	CTGGCGGAGG	57840
	ATCCCGGGTA	CGCCATCGCC	CGGGGCGTTC	CGCTCAACAC	GGACTATTAC	TTCTCGCACC	57900
	TGCTGGGGGC	GGCCTGCGTG	ACGTTCAAGG	CCCTGTTTGG	AAATAACGCC	AAGATCACCG	57960
	AGAGTCTGTT	AAAGAGGTTT	ATTCCCGAGA	CGTGGCACCC	CCCGGACGAC	GTGGCCGCGC	58020

	GGCTCAGGGC	CGCGGGGTTC	GGGCCGGCGG	GGGCCGGCGC	TACGGCGGAG	GAAACTCGTC	58080
	GAATGTTGCA	TAGAGCCTTT	GATACTCTAG	CATGAGCCCC	CCGTGGAAGC	TGATGTCCCG	58140
	CATCTTGCAA	TAAATGTCTG	CGGCCGACAC	GGTCGGAATT	TCCGCGTCCG	CTGGTTTCTC	58200
	TGCGTTGCGT	CTGACCACGA	GCACAAACGT	GCTCTGCCAC	ACGTGGGCGG	CGAACC GGTA	58260
5	GCCGGGGCAC	GCGGTCAGCA	TCCGATCGAT	GAGCCGGTAG	TGCAGGTGGG	CCGACGTGCC	58320
	GGGGAAGATG	ACGTACAGCA	TGTGGCCCCC	GTACGTGGGG	TCCGGGTAAA	AAAGAAACCG	58380
	GGGGTCGCAC	GGCCCCCTC	CGCGCAGGAT	CGTGTGCACG	AAAAAGAGCT	CGGGCTGGCC	58440
	GAGCGTATCG	GCCAGGAGGT	CCTGGAGGGG	GGTGTGTGG	CGGTCCGCCA	GCACGACCAG	58500
	GGAGGCCAGA	AAGGTGCGGT	GCTCAAAGAT	CGTATTGATC	TGCTGCACGA	AGGCCAGGAT	58560
10	GAGGGCCTCG	CGGCTGACGG	TGGCCAGCCG	CCCCTCGCCC	GCGTGCACG	CGGGGCAGCA	58620
	GGCCCCGATC	CCCAGGTAGT	AGCCCATGCC	CGAGAGGGTC	AGGCAGTTGT	CGGCCACGGT	58680
	CTGGTCCAGG	CTGAAGGGGA	GCGACACGGG	GGTCGTCTTC	ACCAGGGGCA	CGGATAGCGA	58740
	GCGCACGATG	GCGATCTCCT	CGGAGGGCGT	CTGGGCGAGG	GCGGCGAAGA	AGCCGCGGTA	58800
	GCGACGGCGC	TCGTGCAGGC	AGAGCTCCAG	CCTGCGCGCG	TGCGACGGCA	GGCTCTTGCG	58860
15	GGAGGCCCGG	CGCTCCACGC	CGGGGTTCCT	GGCGGCGGAA	AAGCGCGACC	GCCGCCGGGT	58920
	CTTGTGCGCG	CCGGGCCCGG	GCCGGGAGCC	GGAGCGACGG	GGGGCGATGT	CATACATAGG	58980
	TACAGAGGGT	GTGCTCCAGG	GACAGGAGAG	AGATCGAGTG	TCGTCTGAGC	AGCGCGCCGG	59040
	CCTCGCGGAC	AAATGTGGCC	AGCGCGGTGG	GCTTCGGCAC	AAATACCTGG	TACGTCTTGA	59100
	AGGTGTAGAT	GAGGGCCCGC	AGGGCTATAC	AGACCCGCCC	CTCGAACTCG	TTGCCGAGG	59160
20	CCAAC'TTGGC	CTTGTGAAGC	TGCAGCTCGT	CGCGATGGTC	GGCGCGGGGG	TGGCCAAACA	59220
	GGACCCAGGG	GTCGACTTCC	ATCTCCGTGA	TGGCGCACAT	GGGATCGCAG	AACATGTGCT	59280
	TGAAGATGGC	CTCGGGGCCC	GCGGCCCGAA	GCAGGCTCAC	GAACCGGCCC	CCGTCCCCCG	59340
	GCTGCGCCTC	GGGGTCCGCC	TCGAGCTGGT	CCACGACCGG	CACTATGCAG	TCGAAGAGGC	59400
	TGGTGTGTGT	CTCCGAGTAG	CGGACGACGG	ACGCCCTCAG	GCGTCGCATG	GCCAGCCAGT	59460
25	AGGCCCCGAC	CAGCAACAGA	TTGCACAGCA	GGCAT'TCCCC	GCCGGTGC GC	CCGCGGCCCC	59520
	GGCCGTGCTT	CAGCACGGTG	GCCATCAGCG	GGCCCAGGTC	CAGGTCGGGC	TGGGCCTTGG	59580
	GCTCGGCGAA	CTGCGCAAAG	CGCGGGGCGG	CGTCGCGCAT	GCGCGCCCCG	CGGTGCGCTT	59640
	CCCAGGACTC	GCTGACCGCG	GCGCGGCGGG	CGTCCGCGGC	GGCGCGCAGC	CGGGGCCCCG	59700
	ACTCCCAGAC	GGCGGGGGTG	CCGGCGAGCA	GCAGCAGGAT	CAGGTCGGCG	TACGCCCACG	59760
30	TCTCCGGCTC	ACCCCCCTGC	GCCAGCGCCC	CGGCGGCGGC	CTCGAACTCC	CCGTTGCGGG	59820
	CGGCGGCGCG	CGTGCAGCAG	CTGTCTCCGC	CCCCGCGCTT	GCCCTCGGTG	CAGTCGAGCA	59880
	GGCGGGCGCA	GTCCTTCCAG	TTCATCAGGG	CGGTGGTGAG	GGAGGGTTGC	GT'TCCCCAGC	59940
	CCCCGCCCCG	CCCCGCCCCG	TCATCGCCCC	CGGAGGCCAG	GGTCCCAGTG	AGGGCCCCGG	60000
	TTGCGGACTG	CGCGAGGAAG	GAATAGTTGG	AGTACTGCAC	CTTGGCGGGC	CCCCGGGAGG	60060
35	GCGTCGGCCT	GGGTGCTTTC	TGGGCGTGCG	GCCC GGGCAC	CCC GCCGTCG	GTCCGGAAGC	60120
	AGCAGTGGAG	AAAGAAATGC	CGGTGGATGT	CGTTGATGGT	CAGGGCGAAG	CGCGCGAAGG	60180
	AGCCGACAAG	GGTCGCCTTC	TTGGTGCGCA	GGAAGTGGTG	GTCCATGACG	TAGACGAACT	60240
	CGAAGGCGGC	CACGAAGATG	CTCGCGGCGC	AGTGGGGCGC	GCCCAGGCAC	TTGGCGCAGA	60300
	GGAACGCGTA	ATCGGCCACC	CACTGGGGCG	AGAGGCGGTA	GGCCTGCTTG	TACAGCTCGA	60360
40	TGGTGCGGCA	GACCAGACAG	GGGCGGTCCA	GCGCGAAGGT	GTCGACGGAC	GCCGCGGCGA	60420
	AGGGCCCCGT	GTCCAAGAGT	CCCTCTGCCG	TGGGGTCTGC	GGGCGGGCCG	CGGGCGGACC	60480
	CGGGCCCCCG	CCCCCCCCGAA	GCCTCGCGCG	CGGCCCGCGG	CGGCCGCGGG	GGGGCGGGCG	60540
	CGACGTCGCT	CTCCACGTCC	TCGTGAGCGG	CGCTCGCGGG	CGGCACGCCT	ACCACGTGAC	60600

	AGGCCGCCAG	GAGCTCGGCG	CACAGGGCCT	CGTTAAGAGC	CAGAAGGTCC	GGATCGAAGG	60660
	CCACATACGG	ACGCTCGAAC	GCGCCCTCCT	TCCAGCTGCT	GCCCCGGCGAC	TCTTCGCGCA	60720
	CGGCGGCGCT	CGACGGCACC	CCCGGGGCGG	ACGTCGCCAT	GGCCGGTCTGA	GCGGGGCGCA	60780
	CGCGTCCGCG	AACGTTACGG	GACGCGATCC	CCGACTGCGC	GCTGCGGTCC	CAGACCTTGG	60840
5	AAAGTCTAGA	CGCGCGCTAC	GTCTCGCGAG	ACGGCGCGGG	GGACGCGGCC	GTCTGGTTCCG	60900
	AGGACATGAC	CCCCGCCGAA	CTAGAGGTTA	TATTCCCGAC	CACGGACGCC	AAGCTGAACT	60960
	ACCTCTCGCG	GACGCAGCGG	CTGGCCTCCC	TCCTGACGTA	CGCCGGGCGCT	ATAAAAGCGC	61020
	CCGACGGCCC	CGCCGCCCCA	CATACGCAGG	ACACCGCGTG	CGTGACCGGC	GAGCTGCTCG	61080
	CCCGAAAGCG	CGAACGGTTC	GCGGCGGTCA	TTAACCGGTT	CCTGGACCTG	CACCAGATCC	61140
10	TGCGGGGCTG	ACGCGCGCTT	CGGCGGGGCA	CCGGCACCGG	GACCGACTTG	TTTTACATAA	61200
	CAGTAGGGGG	TGGGGGAACG	CGCACCCCTG	CCCGGTCGCG	ATGGCGGGGA	TGGGGAAGCC	61260
	CTACGGCGGC	CGCCCGGGGG	ACGCGTTCGA	GGTCTCGTTC	CAGCGCATCA	GGCTCATTTGT	61320
	TCCCACCACG	CTGCGCGGCG	GGGGTGGGGA	GTGCGGCCCC	TACTCGCCAT	CCAACCCGCC	61380
	CTCGAGATGT	GCCTTCCAGT	TCCACGGCCA	GGATGGGTCC	GACGAGGCCT	TCCCAGTCGA	61440
15	GTACGTCCTG	CGGCTCATGA	ACGACTGGGC	CGATGTGCCC	TGCAACCCCT	ACCTGCGCGT	61500
	GCAGAACACC	GGCGTTTCGG	TGCTGTTTCA	GGGGTTTTTT	AACCGGCCCC	ACGGCGCCCC	61560
	GGGGGCGCGG	ATCACGGCGG	AGCAGACCAA	CGTGATTCTG	CACTCCACCG	AGACGACGGG	61620
	ACTGTCCCTC	GGAGACCTGG	ACGACGTCAA	GGGGCGCCTC	GGCCTGACG	CCCGGCCGAT	61680
	GATGGCCAGC	ATGTGGATCA	GCTGCTTTGT	GCGCATGCCC	CGGGTGCAGC	TCGCGTTTCG	61740
20	GTTCATGGGC	CCCAGGACG	CCGTTGCGAC	GCGGCGGATC	CTGTGTGCGG	CCGCCGAGCA	61800
	GGCCCTCGCC	CGTCGCCGCC	GGTCCAGGCG	GTCCAGGAT	GACTACGGGG	CGGTGGCGGT	61860
	GGCGGCGGCG	CACCACTCTT	CCGGAGCGCC	CGGGCCGGGG	GTGCGCCGCT	CGGGCCCGCC	61920
	AGCGCCGCCC	GGACGGGGAC	CGGCCCGTCC	GTGGCATCAG	GCCGTGCAGT	TGTTCCGGGC	61980
	CCCGCGTCCG	GGCCCCCGGG	CGCTTCTGTT	GCTGGTGGCG	GGGCTGTTTC	TGGGGGCCGC	62040
25	TATCTGGTGG	GCGGTTGGCG	CGCGCCTATG	AAAGGGGGCG	AGCCACCGTC	CCGCCGCCCA	62100
	GTGCATCCCA	GACGCCCGCG	AGCCGCACAT	CCCTCCGCT	CCCGCCTCCG	GCCCATTCT	62160
	TACGGCGCGA	CCCAAGGTCC	CGATGGCCGC	CCCGCAGTTT	CACCGCCCCA	GCACCATTAC	62220
	CGCCGACAAC	GTCCGGGCGC	TCGGCATGCG	CGGGCTCGTG	TTGGCCACCA	ACAACGCTCA	62280
	GTTTCATCATG	GATAACAGCT	ACCCGCATCC	GCACGGAACG	CAGGGTGC GG	TGCGAGAGTT	62340
30	TCTTCGCGGG	CAGGCCGCGG	CGCTGACGGA	CCTCGGGGTG	ACCCACGCCA	ACAACACGTT	62400
	CGCCCCGCAG	CCTATGTTTCG	CGGGCGACGC	CGCGCCGAA	TGGCTGCGGC	CCTCGTTTCGG	62460
	TCTTAAGCGC	ACGTATTCCC	CCTTTGTCTG	TCGCGACCCC	AAGACCCCCA	GCACCCCGTG	62520
	AGTCCTCGGC	GGGTCCCTCC	GCGGCCGTCT	CTCGTTGCCC	CCCCTTTCCC	CCTTCCCGGG	62580
	TGGTTCAATA	AAAAACACCA	ACATACGATA	TTGCGGTTTG	ATACGTTTAT	TGGGGGGGTG	62640
35	TAGGGCCCAA	CGATCGGCGA	TTAACAACAC	CAAACAATCG	AGCGCGTCTA	ACCCAGTAAC	62700
	ATGCGCACGT	GATGTAGGCT	GGTCAGCACG	GCGTTGCTGC	GCTGAAACAG	CGCCCTGCGG	62760
	GTCCGCTGCA	GCTGTTGTTG	TATGCGGCGG	CATGCGCGGA	TCAAACCGC	CAGGGCGCTA	62820
	CGACCGGTGC	TTCGTACGTA	GCGTCGCGAC	AAGACGGCAT	TTGCCTGTAC	GGGCAAGGGG	62880
	CCAAATTGCG	AGTGTGGTGA	CTGGAGGTGG	TCGGCGGCCA	ATGGGCCGGG	TGGTTTCGTCG	62940
40	GCGGGGGGCA	AGTGCGGTTC	CGGTGGGAGG	GGGTCGAGCG	CCTCGGTATC	ATCCGAGTCC	63000
	GAGAAACGCA	GGGAGTCTGC	GTCGGAGTGT	TCATCATCGG	AGGAGATGTG	CAGCGTCTGA	63060
	AGCAGCGATG	CGGGTGGGGG	CGCGGAGTCG	ACGTGAAGCG	CGAGAGAGGA	AGCCACGAA	63120
	GTCACAGCGG	ACACTGGGAG	GTGGGTGTTT	GTATGTGTGG	GAGACTCGGG	CGTCGGGACC	63180

	GAGTCTCGGC	TCTGGGGCAG	GGGCGGCTGG	GGCAGGGGCG	GCTGGGGCAG	GGGCGGCTGG	63240
	GGCAGGGGCG	GCTGGGGCAG	GGGCGGCTGG	GGCAGGGGCG	GCTGGGGCAG	GGGCGGCTGG	63300
	GGCAGGGGCG	GCTGGGGCAG	GGGCGGCTGG	GGCAGGGGCG	GCTGGGGCAG	GGGCGGCTGG	63360
	GGCAGGGGCG	GCTGGGGCAC	CGAGCGCGCG	CGGATGCGCG	TCCGCGCGGC	GGGTTTGGTC	63420
5	GCGGGTGACT	GGGGTGGGGG	GCGGCGGGCA	ACCGGGCCTC	CGGGCACGAC	CCAACCGCAC	63480
	AAAGGCTCGC	TCGGGGCAAC	CGGGCCTGGG	GCCAAAGGCG	GGGGCTGGT	CTGGACGGCG	63540
	GAGGTCGGGG	GGGCAAGGCC	CGGAGAAGGC	GGCACTGCCG	CCGCTGCGGC	GGAAACCGCG	63600
	GCCGCGTGGT	CGGCTGGGTC	CCGGGGAGAG	GGGAGGGAGT	TCAACGAGGC	CGAGAGCGAG	63660
	GCGACCGCGG	GGCGCGTGAG	GCGCCGGGGT	GGGCCGGCCG	CGGGGCCCCG	GGGGGGTGTC	63720
10	GGCGAGGGAC	CCGCTGTGTG	CTGGCGGCGG	CCGCGGCGGC	GGTCGTCCCC	GGGGGCGACC	63780
	GCTCCTTCGG	CGGGCGGAGG	CGGGATGGGC	GCGAGCGTGG	GGGCGGGAAA	GGCCCCCGCA	63840
	GCCGAGGCGG	GGCCGGGCGG	AAGGGGCAAA	GCAGAAACCC	AAGCCGGGGG	CGCGGACTCC	63900
	GGGGTGGGCG	GCTGGTCGGG	AGGACGCGCG	GAAGCGGCGA	CCGGGGCGAC	CGGGGCGGGG	63960
	AGTGCCGGCG	GACGCCACCC	CTCGGGGGGG	GCGGAGGCCC	GGGCGCGCGC	CGATTTGGCA	64020
15	CGCGTCCGGC	GGGATCTGCG	CACGCGCGGC	ACGGCGGCGG	AGAAAGCGGC	GGCAGAGCCG	64080
	GAAAAGGCCG	GGGGAGGAAG	CGCGGCATCC	GCGGGGGGAC	TCGGTGTGGG	TGGCGAGGGC	64140
	CGTGCGTCGT	CGCGAGGGGC	CACGGGCACG	CGCCCCGTGT	TTTGTGAGG	CGGGACACTC	64200
	GGTCGTGTTT	CGCGAGCCGT	AGCTGCCGCG	CCGATGGGCC	GCGGTGCGTA	CTGGGACGTG	64260
	GGGACGGAAT	GATCGGTGGC	GGGGGGGGGA	AGAAGGGCCG	GGGCCGGATT	GGGCGTGGGG	64320
20	CCGCCGGCGT	CGTCGGACGC	CAGCTCCTCC	AGGCCGTGGA	TCCAGGCCCA	CATGCGAGGG	64380
	GGGACGGGCT	CGCCGGTGGT	GGCGTCGGTG	AGGAGAGTGG	GGGCGAGGAC	CCCCGGGTCC	64440
	GCCTGCCGCG	CGGGGGGGGC	AGCGGGGTCC	TCGGGACCCG	ATCCGCCATC	CCCCCCCCGA	64500
	AGGTCCCGCG	GGTCGCGGGC	GGCGGTGCGG	GCAGAGGGAC	CTGCCTCGTC	GGCGAGGGGG	64560
	CGCTGGTAAA	CCGGGTGTCC	CGGGAACAGC	TCCCCCGTCA	GGAGGGAGGC	GTCAAGGGGC	64620
25	CGCCCGAGGA	TGGCCCGCGC	GAAGAAGGGG	TCCGCGTCGG	CGGCGCTCGC	CGCGAGAACG	64680
	TCCCCCGCGG	TAGCCACAAA	CGGAAGCTCC	TCGGTGGCCT	CGCTGCCAC	AAACCGCACG	64740
	TCAGGGGGGC	CGGGGGGCTC	CGGGGCTTCC	CACAAGACCG	CGACCGGGGT	CATGGAGATG	64800
	TCCACGAGGA	CCAGGCACGG	GGGCCCCGTC	GCGAGAGGGC	GCTCGGCGAT	GAGCGCCGAC	64860
	AGGCGCGGGA	GCTGCGCCGC	CAGACACGCG	TTTTTCGATC	GGTTGAGATC	GGTGTGGAGG	64920
30	AGGCCGACGG	CCCACGTCTC	GATGTGCGAC	GACACGACGT	CGCGCAGGGC	GGCGTCCGGC	64980
	CCGCCGGGGC	GCGAGTCGAA	GAGCGTCAGG	CACAGTTCCA	GTTCGGAATC	GCGGGAGAAG	65040
	GCCGTGGTGT	TGCGGAGCGC	CACCACGACG	GGCGCGCCGA	GGAGCACCGC	GGCCAGAACC	65100
	AGGTCCATGG	CCGTAACGCG	CGCGGCGGGG	GTGCGGTGGG	TCGCGGCGGC	CAGCACGGCC	65160
	ACGTGCTGGC	CCGTGGGTTC	GTAGAGGGCG	TGGGGGGCCT	CGGGGAGGGA	CGCCTCGCGC	65220
35	CCCCCGCCCG	GGCCGAGCGT	CTGGCCAGAC	TCCAGGCGTG	CGGCCAGGAG	GGCGTCGAAG	65280
	CTGTCTGACT	CGGTGTAGTC	GTCGGGAAAC	ATGCAGGTCC	ACAGCGCGGC	CAAAGCGGCG	65340
	CTCGGCAGAC	ACATGCGCCC	GAGGACGCTC	ACCGCCGCCA	GGGCCTGGGC	CGGACTGAGC	65400
	TTCCCGAGCG	CCGGGGCGTC	CCGGCGCTGG	GTCCCGAGCT	CCAAGGCCGA	GCGCCAGGGC	65460
	GCCAGCGGGT	CGGTTTCGGA	CAGCTTGCCC	CGGCGCCAGT	CGGCCAGCCG	CGTGCCGAAC	65520
40	AGGAGGCCCC	GGGTCGGGGG	GCCTCCGTCC	AAAAACGTCC	GCAACACGCG	GATGCGGGCG	65580
	TCGGGATGCG	GGGTCAGGCG	CTGGACGAAC	AGCATGGACT	CCGCTGCGTC	CTCGAACGCG	65640
	CGTTCGAGGG	TGAGGTGCAT	GTAATCGTGC	TGGCGAACGA	GGTCCAGGCG	CCAGAAGTTG	65700
	TAGATGTGTT	CCGGAACGCC	GGCCACCAGC	GCGACCAGCA	CGTCGTCTTC	GTTGAAGGCG	65760

	ACGCAGTGGC	GCTGGGACCC	CCGGGGGCCC	GGCGGCGGAC	GCGGCGCCGC	CGCTCCGGAC	65820
	GCCAGACCCA	GCTGGGCCCA	GCGACACCCA	AACTCGCGCG	TGAGGGTGGT	GGCGACGAGG	65880
	GCGACGTACA	GCTCGGCCGC	CGCGTCGATC	GAGGCGCCCC	ACGTCGCCTG	GCGATGGCGC	65940
	ACGAAGCGAC	CGAACAGCTG	AAAGTTGGCG	GCCTGGGCGT	CGCTGAGGGC	CAGCTGGAGC	66000
5	CGGTTACGA	CGGTCAGCAC	GTACATGGCC	GTGACCGTCG	GGGCCGATTC	GAGGACGTCC	66060
	GTCGGAAGCG	GGGGCCGCAC	GCAGGCCGCC	TCGGGACGCA	TCAGCAGCGC	GCCGAGTTTG	66120
	TCGGTGACGG	CCGGGAAGCA	TAGCGCGTAC	TGCAGCGGCG	TTCCGTCCGG	GGCCAAAAAG	66180
	CTGGTGCGGA	ACGGCAGATC	CAGAGCGCTG	ACGGCCTCAC	GCAGCACCAG	GGGCCCCGGG	66240
	TCTCCGCCGG	CGCGCAGATA	CGCCTCGCCC	CGGCGGCGCA	GCAGCTGCGG	GTCGACCTCG	66300
10	TGGCCCTCGG	GGGAAGAAGA	GGCCCGGGCG	CGGGCGTCGA	GGGCGCGAAG	ATCAACGAGC	66360
	AGGGGCGCGG	GCGCGGCCGC	CGCGCCCGCG	CCCGTCTGGC	CGCCGGCCCT	GGCGTACGCG	66420
	CTATATAAGC	CCATGCGGCG	TTGGATGAGT	TCCCGCGCGC	CCCGGAATC	CTCCACCGCC	66480
	CACGGGGCCA	GGTCCGCGGC	CGCCGCGTCG	AACTCCGCCA	GCAGGCGCCC	CAGGGCGTCA	66540
	AAGTTCATCT	CCCAGGGCAC	CCTGCGCACC	ACCTCATCCC	GCAGCCGGGC	GCACAGGGCG	66600
15	GTGTGCTTGG	TGACGCGCGC	GCCCAGCTCC	TCCACGGCCT	CCGCGCGCTC	GGCGCCCTTG	66660
	GCGCCAGGA	CGCCCTGGTA	CCTGGCGGAA	AGGCGCTCGT	AGGCCGGCTG	GGCCCGCAGC	66720
	CCCACACCG	TGTTGGTGGT	GTCTGCGAGG	GCGCGCAGCT	GCTCGTGCA	GGCGCGGAAC	66780
	CCCTCGGGGG	ACTTCCAGGC	GCCCCCCC	ACGCGGCCAA	AGCGACCCCA	GACCTCGTCC	66840
	CACTCCGCC	CGGCCCTC	CAGGGACCTC	CGCAGGGCGT	CGACGCGGCG	CCGAGTATCA	66900
20	AAGAGCGCCC	CCAGGCGGCC	GGCGTGCCGC	GCCAGGGGGC	CGGGGCCGTC	GCCGCGGGCG	66960
	GCGCTTAGCG	GGTGCGTCTC	GAAGGTGCGC	TGGGCGTGCT	CTAGCCAGAT	AACCGCGGGC	67020
	ACGTCGAGCT	CGCGCGTTTT	CTCGGTCTGA	TCCAACAGAA	CCTCGACCTG	GTCGGCGATC	67080
	TCCGCCACCG	AGCGCGCCTG	GTCGAGCGTC	TTGGCCACGG	TCGCCGGGAC	GGCGACCACC	67140
	TTCAGCATGG	TCTTGAGGTT	GGCCAGGCC	TCGGCCTCGA	TCTGGGCCCC	GCGCTCGCGC	67200
25	GCGGCCAGCG	CCTCCCGCAG	GCCCGCCATG	ACCCGCTCGG	TGGCCTCCGC	GCGCTGCTGT	67260
	TTGGCGCGCA	CCACTGCGTC	CTTGGTCTCG	GCCGTGTCT	GCCGGGTCAC	GAAGGCGACA	67320
	TACTCGGCGT	ACGCCGTGTT	CTTACGGGG	CTCTGGTCCA	CGCGCTCCAA	CGCCGCCGCG	67380
	CACGCGACCA	GCGCGTCCTC	GCTGGGACAC	GGCAGGGTGA	CCCCGGTCCG	GACCAGCTCC	67440
	GCGGTGGCCT	CCGGGTCATT	CCGGGCCGCG	GATATCTGCT	CCGCGGCGGC	CGCCAGGTCC	67500
30	AGGGGCACGC	CGCCGAGCGC	CCGGTGACAG	TCGGCCCGGA	TGGCGTCCAG	GCGATCGCGG	67560
	AGCTCCACGT	AGTCGGCGTA	GCCATGTTGG	AAGAACGGCA	CGTACCGGCG	CAGGCCGGGC	67620
	ACGCTCGTCA	TGTCGTCCGC	CAGGCGCCCC	ACGGCCTCGT	GGTAGTCGAT	AAACCCGTCTG	67680
	CCCGCCTGGG	CCATTTCAG	GAGCCCCCTC	GCGATGCGCA	GCAGCCGCGC	CAGGGGCTCG	67740
	GCGTCGACCC	GAAACATGTC	GGCGTAGGTT	TCGGCGGCGG	CGTGGAACGC	CGCGCTCCAG	67800
35	CCGAGGCGGT	GGATGGCGGC	GAGCGGGGGG	AGCATGGGGT	GGCGCTGGTT	CTCGGGGGTG	67860
	TAGGGGTAA	ACGCGAAGGC	CGTATCCAGG	GCGAGGGTGA	CCGCCTCGGC	GTTGGCCGCG	67920
	AGCGCCTGCT	CGGCGCGCTT	GCGGAAGTCC	CGGGGGTTGT	AGCCGTGCGT	GCCCCCAGC	67980
	GCCTGCAGGC	GGCGCAGCTC	GACCACGTCG	AACTCGGCGC	GGTTCGAC	GCGGTCCAGC	68040
	GCCGCCTCGA	CGCCGGCGGC	CCAGCGCTCG	CTGCTGCCCC	GGGCGCGCTG	GGCCGCCATC	68100
40	TTCCGCGTCA	GGTCGGCGAC	GGCGGCCTCA	AGTTCGTCGG	CGCGGCGTCG	CGTGGCGCCG	68160
	ATGACCTTGC	CCAGCTCCTG	CAGGCGCGCG	CCGCTGGGGG	AATGGTCCCC	GGCCGTCCCT	68220
	TCGGCGTGCA	GCAGGCCCCC	GAACCCAGCC	TCGTGCCCCG	CGAGGCTTTC	CCGAGCAGCG	68280
	GTCGTGCGCG	GGGCCGCGGC	ATCGATGAGG	GCGGCATGGT	CCCCCTCCGG	CTGGGCGCAG	68340

	GCCCCGCGCG	CCTGGACTAC	CAGGTCGGCG	GCCGCCGACC	CCAGGGTCGT	GAGCTCGTCG	68400
	ATGGCCCCC	GCGCCTCCAG	GGCCAGCCGA	GTCGCTTTA	CATACCCCGC	GGCGCTATCG	68460
	GCCAGCGCCG	CGAGGAAGGA	CAGGGGEGAG	GCCGGGTCGC	GGGCGGCCCG	GCCCAGGGCC	68520
	GACACCGCGT	CCGCCAGGGC	GCCATGCGCC	CGCACGGCCG	CGTCCACCGT	CGCCGCGGGA	68580
5	CTTGCCGTCG	CGACGGCGGC	GCTCCCGGCG	TTGATGGCGT	TTGACACGGC	TTTGGCGATT	68640
	GTGGGGGCGT	GATCGGAAAA	GAAGTGCACG	AGGACCGGCG	TCTCGGGGGC	GTCGGCGAAC	68700
	AGGGTCTTCA	GCACCACCAC	GAAGGCGGGA	TGCAGGCCCG	CCAGAGCCGT	CGCGGTATCC	68760
	GGGGTCGGGT	GTTCACGGGC	CTCCCGGTAC	TGCCCCAGCA	GCCCCACAG	GTCCGCCCCG	68820
	AGCGCCGCCG	TGACTTCCGG	GGGGGGGGCC	CGGACGGCAT	CGGCCAGGTC	GGTCCACCCC	68880
10	GCGGGCAGGG	AGGCCCCGAG	GGTCGCCAGC	ACGGCCGGAC	ACGCCTTTAG	CCCCACAAAG	68940
	TCCGGGAGGG	GCCGCAGGAC	CCCTTGGAGT	TTGTGCAGGA	ACTTCTCCCG	GGCGTCGTGG	69000
	GCCACCTTGG	CGCGCTCCCG	CGCGTCGTG	AGCATCGCCT	CCAGGGCGTG	GGCGCGCTCC	69060
	CGAAGCCGGG	AGCGCGCCTC	CGGAGCGAGC	TCCGCCGTCA	TCTTGGCCGC	CTCCATGGCC	69120
	CTCGCCTGCC	GCAGCGCGTC	TTCGGCCATG	CGCGTGGCCT	CGGGGGACAG	CCCGCCCCCG	69180
15	TCGACGTACG	GCGCGGGGCC	GGTCGCCGGG	ACGAAGGCCG	CGTCGCTGTC	CAGCTGCTGC	69240
	GCGAGCGCCG	CGTCGAGGGC	GTCGAAGCGC	TGCAGTTCGG	CCAGCCCCGA	GCTGCGCCGC	69300
	GCCTGCTGGT	CGTTGATGCC	GTGGATGCTG	CGCGCCAGCT	CTTCCAGGGG	CTTGCCTTCG	69360
	ATGAGCCCCCT	GGGTCGCGGC	GTCGGTCAGG	ACCAGAGGCC	AGGCCGCCAG	GTCCTCGGGG	69420
	GCATCTAGGG	TCTGGCCCCG	CTGGAGCAGG	TCCCGCAGCA	GGATGGCCTG	GGGGCTGGTG	69480
20	GCGAGGGGGG	GCGGGGGGGG	GAGCGCGGCG	CGCTGAGCGA	CGTCCCGCGT	GTGTTGGTCA	69540
	AAGGCCGGTA	GCGATTCCAG	CAACTGGACC	ATGGGCACGA	CCGCGGCCGA	GGCCACGTGA	69600
	AACCGACAGT	CGTGGCTGTC	GCTGGCCTGC	AGGGCCTTCG	CGCTGTATAC	GGCTCCCCCG	69660
	TGGAAGTACT	CCTTGATCGC	GCTCTCGATC	GCCCCGCGGG	CCTGGATCCG	CACGTCTCTC	69720
	AGCCGCGCCT	GGATGGCCTC	GGGGCCCCAG	GCGGGCGGGC	ACGGGGCCCT	GCCGCGGGCG	69780
25	CCCCGGGCGG	CGGGCACGGG	CATCACGGTC	AGGGGCCCCG	CGCGCTGCGA	GACCGAGTCG	69840
	ACCCCGCGGG	CGAGGGCGTC	TAAGGCCTCG	CGCATCTCGC	GGGCCTCCGC	CTCGACCCGC	69900
	ATCTCTTCGC	CCCCGGCAAA	CTGGGCCAGC	GCCTGGATCC	GATGGAGAAG	CGGCTCCGGG	69960
	TGGTCGGGGG	TGGCGGGGGC	GAACAGGGTG	TTCGGGTGGG	CGCGCGAGCG	CTCCAGGAGC	70020
	CACTCTCCGA	GGCGTGCGTA	CAGATTGGCC	GGCGGGGCGG	CGCGCAGCTG	CAGATCCAGG	70080
30	TCCGCGAGGT	CCCCGTAAAA	GGCGTCCGTC	TCCCGAATAA	CGTCCCTGGC	GACCAGGACC	70140
	AGCTTAGCGA	GGGCCAGGCG	CCCGATCTGC	GAATTTTCGT	CCAGCACGTG	CTGGATGAGG	70200
	GGCCGGTGGG	CGGCCACGTC	CGCCAGGCTC	ATGCGCGTGG	ACGCCAGGAA	GTCCCCGACG	70260
	GCCGTTTTCG	GGGGCAGCAT	GCGCAGGGTG	AAGTCCAGCA	GGGCCGCGGC	CGGGCCGGCC	70320
	ACCCCGGCCT	GCGTATGCGT	GCGGGCCCCG	TTCTCGATCA	AAAAGGCGAG	GACGCGCTCA	70380
35	AAGAAGAAGA	TGACGCAGAG	CTCCAACAGC	CCCGGGTGCG	CCGGGTACGG	CGACCGCAGG	70440
	GCGTTGATGG	TGAGCTGCGA	ACACGCGGCC	ACCTCGCGGG	CCAGGGCGGC	ATCGCGCGCC	70500
	GCGAGCCGGA	CCGCCGTGGC	GGCCACATTC	GGGTGGACCT	CGAACAGCTG	CGCCAGGTTCG	70560
	GCGCCGGGGG	GCTCCGGGGG	GCGGCGGGCC	CCCAGCGTCT	CGAGCACGGA	CGGCGACGAC	70620
	GGGTCGCGCG	GCCCGTCATC	GCCGCTCCCG	TGCCCGGACT	GCGGGGGGGT	ATCCGGTGCG	70680
40	GGAGGGACCG	TGGCGGCTAT	GGGCGTCGGG	GAGGAGGCGG	GGACCTCGGC	GGCGACGGGG	70740
	GCCCTCTTCT	TGGGCGCGGA	CTTCTTCTTG	GCCTTGGCGG	GCGGGGCCCTT	GGGGGCGGGC	70800
	CTCTCGCCCG	AGGTCAGATC	CTCCACGCTG	GACGGTGGGG	TCCAGGTGGG	CCGGCGGCGC	70860
	TTGGGCAAGC	CGGTAGAATA	GCGCGCCCCG	TGGCGACCCA	CCGGCACTGC	CCCCACCTCC	70920



	AGGACCCGCA	GGTCTCGGC	TTCTTCGGCC	GCGTCCCCGG	CGGGTGTCTG	CGGGGGCGGG	70980
	GCGGCGTGCG	GTGGACCCGA	GGCCGCGGCG	TCCGGGGCCG	AGGGCTTTGC	GGGCGGGGTC	71040
	CCCTCCAGGG	CTGCTGCCCA	CACATCATCG	GGGGGGCGGT	TTGGGTGCCC	CGCCTGCCGT	71100
	GTGTTGGGTG	GGCCCCAGGC	CCCCCGGGGG	GCCTCGGGGG	GCCGGTCGGC	CCGAGGGGTC	71160
5	TGGACGTGGG	TGGGCGCGGG	GAGCGCGGGG	ACGACCGGGC	CCGAGCCTTC	TCCGTCCCCC	71220
	CTGGGGACCA	CACCGACAAA	GAGCGCCCCCT	AGCCCCCGA	TCTCGCCCCG	CAGGGGGTGG	71280
	GTGATGGCCA	CGCGCCGCTC	GACGAACGGT	TCGTCTTGCA	GGTAAGTCTC	GCTGGCCCCG	71340
	TAGAGGTGCA	GGGCCGCGGC	GGTCAGGTCC	GCCGGCGCCA	CGGCCCCCGG	GCCGGAGGGC	71400
	ACAAAAACA	CCATGGCGCC	CGCCACCGC	ACCTTGGGGC	GGTCGTGGGC	GTAATACGTC	71460
10	AGGTACGGGT	ACACGTCGCC	CGCCCGCACC	TTGGCGATAA	ACGCGGGCGT	TCCCGCGGGC	71520
	AGGCCGTGCG	GGTCAAACAG	ATAGGCCGTG	TCGCCGTCCC	GGTAGAGCCC	CATGCCCCAGG	71580
	GGGCCGATGG	TCAGGAGCGT	GTAGGACAGC	GGCCGCATGG	CCCAGGGGCC	GGCGAAGAAC	71640
	GTGTGCGCGG	GGCATTGCGT	CTCCAGCAGC	CCCGCCGTGG	GCTCCCCGAA	GAAGCCCACC	71700
	TCGCCGTACA	CCCGCGAAAA	CACGCAACGC	AGGCCGCCGC	GCGCCGCCGG	GTAATCCAGG	71760
15	AAGTTGGGGA	GCTCGATAAT	GGAACACATG	CGCGGCGGCC	CGGAGCCCCG	GGCCGCGCGC	71820
	GTCCACTCGC	CCCCCTCCAC	CAGACATCCC	TCAATGGCCT	CCGCGGACAG	CACGTGCGCG	71880
	GGCCCCACGT	CGAAAAGAAG	ACTGAGAAAC	GACAGGGACG	AGCGCATGCA	CGATACCGAC	71940
	CCCCCGGCT	CCAGATCGGT	CGCGAACTGG	TTCCGAACAC	CGGTGACCAC	GATATCGCGA	72000
	TCCCCCTGGC	GCTTCATCGT	GGGGTGAGGT	AGCGCGGCCG	GAATCATGTG	TGCCGCGCCC	72060
20	GCCACGAGCG	GGGCTGT'TT	ATGGGCGGGG	CGTCCCGATG	AGTACTGTTG	TTTCCGCCGC	72120
	CCGAACCCCC	CCCCCCCATC	AACCGCCTGT	TCGTCCCCCT	AACCACACAC	CCGGTATCGC	72180
	GTGTGTGGTT	TCCCGGAAG	CCACATCCCA	CCCCATGAAG	TTTTGCCCTT	TTTTTCCGTC	72240
	CCGCACTACG	CCACCTTTCC	ACCCCCCCCC	CCCAAAAAAA	AAAAACAAC	AACCAACTCC	72300
	CAGATGGATG	GGTGCGATAA	TAAAGCTTTA	TTATTGTTTA	ACCAAAGGCG	AGTCCTACGG	72360
25	GTGTACCGGT	GGTGCTCTCT	GCGGCGTCAT	CTCGTCGTCC	TCCACGGGGG	TGTTGGGCCA	72420
	AGGGACCGTC	TCGCGGCCCG	CCGGGCGCGT	CGACGGCGCG	CGGGCCTGCG	TGTCCTGTGG	72480
	GCCGGGTGTC	GTGGGTTCGG	GGGTGCTACC	GCCGGCATCT	TGGGCTCCA	GGTCCCCGGG	72540
	GGCCTCCGGG	CCGGCGGAAG	GCCGAAACGC	CGAGGCGCGA	AACACGCCGT	CGGTGACCTG	72600
	CAGGAGCTCG	TTTATTAATA	GCCAGTCCAT	GCTCAGCGTA	GCGGCCAGCC	CCTGGGGAGA	72660
30	CAGGTCCACG	GAGTCCGGAA	CCACCGTCGG	CTGACCCAGG	GGCCCCAGGC	TGTAGTCCCC	72720
	CCAGGCCCCC	AGGTCATGAC	GGTTCGTGAG	CACGACGAGG	TCTGCGGCCG	GGCTGGGGGG	72780
	CGCGTCTCG	GTCGCGTGGG	CCATCACCTC	CTGAATGGCT	GCGGTGCGCT	GATCGGCCGA	72840
	GCTGGCGAAG	GGCGCCACGA	CCAGCGCGCG	CTCCGTCTGC	AGGCCCTTCC	ACGTGTCTGT	72900
	GAGTTCCTGA	ACGAACTCGG	CCACCCGCTC	GGGGCCCCGTG	GCCGCGCGCG	CGGCCTGATA	72960
35	GCCGGCCGAG	AGGCGCCGCC	AGCGCGCCAG	GAAC TGACTC	ATGTAACAGA	ACCCGGGGAC	73020
	CTGGTCCCCC	GACATCAACT	TTGACGCCCT	GGCGTGGATG	CCCGACACGA	TGGCCAGGAA	73080
	CCCGTGGATT	TCCCGCCGCA	CGACGGCCAG	CACGTTACCC	TCGTGCGAGA	CCTGGGGCCG	73140
	CAGCTCGTCG	CATACCCCGA	GGTGCGCCGT	CGTCTCGGTG	ACGACGGACC	GCAGCCCCGC	73200
	GAGGGACGCG	ACCAGCGCGC	GCTTGGCGTC	GTGATACATG	CCGCAGTACT	GGCTCACCGC	73260
40	GTCGCCCATG	GCCTCGGGGG	GCCAGGGGCC	CAGGCGCTCG	TGGGCGTCTG	CGACCACGGC	73320
	GTACAGGCGG	TGCCCCGTGCG	TCTCGAACCG	GCACTCAAAG	AAGGCGGCGA	GCGTGCGCAT	73380
	GTGCAGCCGC	AGCAGCACGA	TCGCGTCCCT	CAGCTGGCGG	ACCAGGGGGT	CGGCGCGCTC	73440
	GGCGAGCTCC	TGCAGCACCC	CCCCGGGCCG	CAGGGCGTAC	ATGCTGATCA	GCAGCAGGCT	73500

	GCTGCCCACC	TCGGGAGGCT	GGGGGGGAGG	CAGCTGGACC	GCGGGCCGCA	GCTGCTCGAC	73560
	GGCCCCCTTG	GCGATCACGT	ACAGCTCGCG	CAGCAGCTGC	TCGATGTTGT	CGGCCATCTG	73620
	CATCGTGGGC	CCGACGCCGG	CCCGGGTGGC	CGGTTCGAGG	AGGGTGATCA	GCGCGCCCAA	73680
	TTTTGTGCGG	TGCCCTCGA	CGGTGGGGAG	ATAGCCCAGG	CCGAAGTCGC	GCGCCCAGGC	73740
5	CAGCACCCGC	AGGGCAAAC	CGATGGGGCG	GGGCAGGTAG	GCAGCGTTGC	ACGTGGCCCT	73800
	CAGCGCGTCC	CCGACCACCA	GGGCCAGCAC	GTAAGGGACG	AACCCCGGGT	CGGCGAGGAC	73860
	GTTGGGGTGG	ATGCCCTCCA	GGGCCGGGAA	GCGGATCTTG	GTGGCCGCGG	CCAGGTGAAC	73920
	CGAGGGGGCG	TGGCTAGGCG	GCCCCAGGGG	GAGCAGCGCG	GACAGCGGCG	TGGCCGGGGT	73980
	GGTGGGGGTC	AGGTCCCAGT	GGGTCTGGCC	GTACACGTCG	AGCCAGATGA	GCGCCGTCTC	74040
10	GCGCAGGAGG	CTGGGCTGGC	CGGCGCTGAA	GCGGCGCTCG	GCCGTCTCAA	ACTCCCCCAC	74100
	GAGCGTGC	CGCAGGCTCG	CCAGGTGTTC	CGTCGGCAGC	GCCGGGCCCCA	TGATGCGCGC	74160
	CAGCGTCTGG	CTGAGGACGC	CGCCCCGACAG	GCCGACCGCC	TCACAGAGCC	GCCCGTGC	74220
	GTGCTCGCTG	GCGCCCTGGA	TCCGCCGGAA	CGTTTTTCACG	TAGCCGGCGT	AGTGCCCGTA	74280
	CTCCCGCGCG	AGCCCGAACA	CGTTCGCCCC	CGCAAGGGCA	ATGCACCCAA	AGAGCTGCTG	74340
15	GATCTCGCTG	AGCCCGTGGC	CGGGGGGCGT	CCGCGCGGGC	ACCCCGCCA	CCAAAAACCC	74400
	CTCCAGGGCC	GATATGTACT	GGGTGCAGTG	CGCGGGCGTG	AACCCCGCGT	CGGTAAGCGT	74460
	GTTGATCACC	ACGGAGGGCG	AGTTGCTGTT	CTGGACCAA	GCCACGTCT	GCTGCAGCAG	74520
	CGCGAGGAGC	CGTTGCTGGG	CCCCGGCGGA	GGGCGGCTCC	CCTAGCTGCA	GCAGGCCGGT	74580
	GACGGCCGGA	CGGAAGATGG	CCAGCGCCGA	CGCACTCAGA	AACGGCACGT	CGGGGTCGAA	74640
20	GACGGCCGCG	TCCGTCCGCA	CGCGCGCCAT	CAGCGTCCCC	GGGGGCGCGC	ACGCCGACCG	74700
	CGGGCTGACG	CGGCTTAGGG	CGGTGACAC	GCGCACCTCC	TCGCGACTGC	GAACCATTTT	74760
	GGTGGCCTCG	AGGGGCGGGA	TCATGATAGC	CGGGTCGATC	TCCCGCACCG	TGTGCTGAAA	74820
	CTGGGCCAGC	AGCGGCGGCG	GGACCACCGC	GCCCCGATCG	GGGGTCGTCA	GGTACTCGTC	74880
	CACCAGCGCC	AGCGTAAACA	GGGCCCCGCGT	GAGGGGGGTC	AGGGCGGCGT	CGTCGATGCG	74940
25	CTGTAGGTGC	GCCGAGAACA	GCGTCACCCA	ATTGCTGACC	AGGGCCAAGA	ACCGGAGACC	75000
	CTCTTGACAG	ATCGGGGACG	GGAAGAGCAG	GCTGTACGCC	GGGGTGGTCA	GGTTGGCGCC	75060
	GGGTTGCCCC	AGGGGAACCG	GGGACATCTT	AAGCGACATC	TCCCGAGGG	CCTCCAGGGA	75120
	GGTCCGCGGG	TTCATGGCCA	GGCAGCTCTG	GGTGACGGTC	CGCCAGCGGT	CGATCCACTC	75180
	CACGGCACAC	TGGCGGACGC	GCACCGGCCC	CAGGGCCGCC	GTGGTGCGCA	GCCCGGCGGC	75240
30	CTCCAGCGCG	TGGGTCGTGT	CGGAGCCGGT	GATCGCCAGG	ACCGTGTCTT	TGATGACGTC	75300
	CATCTCCCGG	AAGGCCGCCT	CGGGGGTCTC	GGGGAGCGCC	ACCGCCATGC	GGTGACACCAG	75360
	CAGCCCGGGG	AGGTCTCTCG	CCAAGAGCGC	CGTCTCCGGA	AGCCCGTGGG	CCCGGTGCAA	75420
	GGCGCACAGT	TGCTCCAGGA	GCGGGTGCCA	GCACGCCCCG	GCCTCCGCGG	GGCCGACCGC	75480
	CGCGCCCGAC	AACAGAAACG	CCGCCGTGGC	GGCGCGCAGT	TTGGCCGCGG	ACAGAAACGC	75540
35	CGGCTCGTCC	GCGCTGCCCC	CCGGCTCGCT	CGAGGGGGAG	GGCGGCCGGC	GGAGGTTGGT	75600
	CAGGCTCCCC	AACAGGACCT	GCAACGGTCC	GTTTGGGGGT	GGAGCGGACG	GGGGGGTCAT	75660
	GCCGGCGGGC	GCCGGGACCT	GGAGCGCGCT	GTCCGACATG	GCGACCGGCG	TGCGCGCTCG	75720
	GCGACGCGGC	GCGGAGACCG	CGGGCCCAAA	CGGGAATGAC	TGCCCGCGCC	CTATACGGAG	75780
	GGGCTAAGTA	TGCCCCGGGG	ACCCTTCGAA	ACCCCGGGCG	TGTCGCAAGT	ACGCCCGCAA	75840
40	GGCGCGCGGT	GTTATACGGC	GCGTTATGTC	CCGGCATTC	GTTCTGTGGT	TCGGGCCCGG	75900
	GTGCTGTGCG	GTGGGAGTGT	GTGTGTGTGG	GGGGGGGGCG	GCGCGACGGC	GGCCCGGACC	75960
	AAGTGTATCG	CGGCCGTTC	GTGGGGCGGC	CCAACAGGCC	CTTTAAACAT	TTGCGTATGC	76020
	ACCGGCCAG	CCAGTCGGAC	ACCGGAACCC	ACCAGAGGCG	GAAGCCGCCT	TCGCCCGTGA	76080

	GGGTGCGTGT	GTTTTCTGGT	GGCGTGTTTT	TCCTTTCCGC	CCTCCTCCCT	CCCCACCTCC	76140
	ACCACCCCC	CCCCACAAC	CGCCCCGTTG	CGATCGGCGG	GAAAACCATG	AAAACCAAGC	76200
	CACTCCCGAC	AGCCCCGATG	GCGTGGGCGG	AGAGTGCCGT	GGAAACCACC	ACCAGCCCGC	76260
	GCGAGCTCGC	GGGCCACGCC	CCGCTCCGGC	GCGTCTGCG	CCCGCCCATC	GCTCGCCGCG	76320
5	ACGGCCCGGT	GCTTTTGGGG	GACAGGGCCC	CCAGGAGGAC	GGCCAGTACG	ATGTGGCTGC	76380
	TGGGGATCGA	CCCCGCGGAG	TCGTCTCCGG	GAACGCGCGC	TACCCGAGAC	GATACCGAGC	76440
	AGGCCGTGGA	CAAGATCCTC	AGGGGAGCCC	GGCGCGCGGG	AGGGCTGACC	GTCCCCGGCG	76500
	CCCCCGCTA	TCACCTGACC	CGCCAGGTAA	CCCTGACGGA	TCTCTGCCAA	CCAAACGCGG	76560
	AGCGGGCCGG	GGCGCTCCTT	TTGGCCCTGC	GGCACCCAC	CGACCTCCCC	CACCTGGCCC	76620
10	GCCATCGGGC	TCCGCCCGGC	CGGCAGACCG	AGCGACTGGC	CGAGGCCTGG	GGCCAGCTCC	76680
	TGGAGGCCTC	CGCCCTGGGG	TCCGGGCGGG	CCGAGAGCGG	CTGCGCGCGC	GCGGGCCTTG	76740
	TGTCGTTTAA	CTTCTGCTG	GCCGCGTGCG	CCGCCGCTA	CGATGCGCGC	GACGCCGCCG	76800
	AGGCGGTCCG	GGCCACATC	ACGACCAACT	ACGGCGGGAC	GCGGGCCGGG	GCGCGGCTGG	76860
	ACCGGTTTTT	CGAATGCCTG	CGCGCCATGG	TCCACACGCA	CGTGTTTCCC	CACGAGGTCA	76920
15	TGCGGTTTTT	CGGGGGGCTA	GTGTGCTGGG	TCACACAGGA	CGAGCTGGCT	AGCGTCACCG	76980
	CCGCTGCGAG	CGGACCCAG	GAGGCCACAC	ACACCGGCA	CCCGGGCAGG	CCCTGTTCCG	77040
	CCGTTACCAT	CCCGGCCTGC	GCCTTCGTGG	ACCTGGACGC	CGAGCTGTGC	CTGGGGGGCC	77100
	CTGGGGCGGC	GTTCCTGTAC	TTGGTCTTCA	CCTACCGACA	GTGCCGGGAC	CAGGAGCTCT	77160
	GTTGCGTGTA	CGTGGTCAAG	AGCCAGCTCC	CCCCGCGCGG	ACTGGAGGCG	GCCCTCGAGC	77220
20	GGCTGTTCGG	GCGCCTCCGG	ATAACCAACA	CGATTACCGG	GGCCGAGGAC	ATGACGCCCC	77280
	CTCCCCCGAA	CCGAAACGTT	GACTTTCCGC	TCGCCGTCTT	GGCCGCGAGC	TCGCAATCCC	77340
	CGCGGTGCTC	GGCGAGCCAA	GTCACGAACC	CCAGTTTGT	CGACAGGCTG	TACCGCTGGC	77400
	AGCCGGATCT	GCGGGGGCGC	CCTACCGCAC	GCACCTGCAC	ATACGCCGCC	TTCGCAGAGC	77460
	TGGGTGTCAT	GCCAGACAAC	AGCCCCCGCT	GTCTGCACCG	CACCGAGCGG	TTTGGGGCGG	77520
25	TCGGCGTTCC	GGTTGTCATC	CTGGAGGGCG	TGGTGTGGCG	CCCCGGCGGG	TGGCGGGCCT	77580
	GCGCGTGATC	GTCTATTGAC	GACGGCCGCC	CAACCCGAGC	GACCTTCCCC	TCCCACCTTC	77640
	CCCCCCCCC	CTCCTACACA	CCAACTCCGC	CCTCGCCGTC	TTGGCCGTGC	GCGGCCCCGT	77700
	GCGTCCGTCT	CAATAAAGCC	AGGTTAAATC	CGTGACGTGG	TGTGTTTGGC	GTGTGTCTCT	77760
	GAAATGGCGG	AAACCGACAT	GCAAATGGGA	TTATGGACA	CGTTACACCC	CCCTGACTCA	77820
30	GGAGATAGGC	ATATCCTCCT	TAGATTGACT	CAGCACACGA	TCGCACCCCA	CCCCTGTGTG	77880
	CCGGGGATAA	AAGCCAACGC	GGGCGGTCTG	GGTTACCACA	ACAGGTGGGT	GCTTCGGGGA	77940
	CTTGACGGTC	GCCACTCTCC	TGCGAGCCCT	CACGTCTTCG	CCCACCGATT	CCTGTTGCGT	78000
	TCCTGTGCGG	CGGTGCTGTC	CTGTGACAG	ATTGTTGGCG	ACTGCCCGGG	TGATTGCTCG	78060
	GCCGGTGCGT	CCTTTCGGTC	GTACCGCCCA	CCCCGCCTCC	CACGGGCCCC	CCGCTGTTTC	78120
35	CGTTCATCGC	GTCCGAGCCA	CCGTCACCTT	GGTTCCAATG	GCCAACCGCC	CTGCCGCATC	78180
	CGCCCTCGCC	GGAGCGCGGT	CTCCGTCCGA	ACGACAGGAA	CCCCGGGAGC	CCGAGGTCGC	78240
	CCCCCTGGC	GGCGACCACG	TGTTTTGCAG	GAAAGTCAGC	GGCGTGATGG	TGCTTTCCAG	78300
	CGATCCCCC	GGCCCCGCGG	CCTACCGCAT	TAGCGACAGC	AGCTTTGTTC	AATGCGGCTC	78360
	CAACTGCAGT	ATGATAATCG	ACGGAGACGT	GGCGCGCGGT	CATTTGCGTG	ACCTCGAGGG	78420
40	CGCTACGTCC	ACCGGCGCCT	TCGTGCGGAT	CTCAAACGTC	GCAGCCGGCG	GGGATGGCCG	78480
	AACCGCCGTC	GTGGCGCTCG	GCGGAACCTC	GGGCCCCGTC	GCGACTACAT	CCGTGGGGAC	78540
	CCAGACGTCC	GGGGAGTTCC	TCCACGGGAA	CCCAAGGACC	CCCGAACCCC	AAGGACCCCA	78600
	GGCTGTCCCC	CCGCCCCCTC	CTCCCCCCTT	TCCATGGGGC	CACGAGTGCT	GCGCCCGTCG	78660

	CGATGCCAGG	GGCGGCGCCG	AGAAGGACGT	CGGGGCGCGG	GAGTCATGGT	CAGACGGCCC	78720
	GTCGTCCGAC	TCCGAAACGG	AGGACTCGGA	CTCCTCGGAC	GAGGATACGG	GCTCGGGTTC	78780
	GGAGACGCTG	TCTCGATCCT	CTTCGATCTG	GGCCGCAGGG	GCGACTGACG	ACGATGACAG	78840
	CGACTCCGAC	TCGCGGTCGG	ACGACTCCGT	GCAGCCCCGAC	GTTGTCTGTC	GTCGCAGATG	78900
5	GAGCGACGGC	CCTGCCCCCG	TGGCCTTTCC	CAAGCCCCGG	CGCCCCGGCG	ACTCCCCCGG	78960
	AAACCCCGGC	CTGGGCGCCG	GCACCGGGCC	GGGCTCCGCG	ACGGACCCGC	GCGCGTCGGC	79020
	CGACTCCGAT	TCCGCGGCCC	ACGCCCGCCG	ACCCCAGGCG	GACGTGGCGC	CGGTTCTGGA	79080
	CAGCCAGCCC	ACTGTGGGAA	CGGACCCCGG	CTACCCAGTC	CCCCTAGAAC	TCACGCCCCG	79140
	GAACGCGGAG	GCGGTGGCGC	GGTTTCTGGG	GGACGCGGTC	GACCGCGAGC	CCGCGCTCAT	79200
10	GCTGGAGTAC	TTCTGTCTGGT	GCGCCCGCGA	GGAGAGCAAG	CGCGTGCCCC	CACGAACCTT	79260
	CGGCAGCGCC	CCCCGCCCTCA	CGGAGGACGA	CTTTGGGCTC	CTGAACTACG	CGCTCGCTGA	79320
	GATGCGACGC	CTGTGCCTGG	ACCTTCCCCC	GGTCCCCCCC	AACGCATACA	CGCCCTATCA	79380
	TCTGAGGGAG	TATGCGACGC	GGCTGGTTAA	CGGGTTCAAA	CCCCTGGTGC	GGCGGTCCGC	79440
	CCGCCTGTAT	CGCATCCTGG	GGATTCTGGT	TCACCTGCGC	ATCCGTACCC	GGGAGGCCCTC	79500
15	CTTTGAGGAA	TGGATGCGCT	CCAAGGAGGT	GGACCTGGAC	TTCGGGCTGA	CGGAAAGGCT	79560
	TCGGAACAC	GAGGCCCAGC	TAATGATCCT	GGCCCAGGCC	CTGAACCCCT	ACGACTGTCT	79620
	GATCCACAGC	ACCCCGAACA	CGCTCGTCGA	GCGGGGGCTG	CAGTCGGCGC	TGAAGTACGA	79680
	AGAGTTTAC	CTCAAGCGCT	TCGGCGGGCA	CTACATGGAG	TCCGTCTTCC	AGATGTACAC	79740
	CCGCATCGCC	GGGTTCCTGG	CGTGCCGGGC	GACCCGCGGC	ATGCGCCACA	TCGCCCTGGG	79800
20	GCGACAGGGG	TCGTGGTGGG	AAATGTTCAA	GTTCTTTTTC	CACCGCCTCT	ACGACCACCA	79860
	GATCGTGCCG	TCCACCCCCG	CCATGCTGAA	CCTCGGAACC	CGCAACTACT	ACACGTCCAG	79920
	CTGTACCTG	GTAAACCCCC	AGGCCACCAC	TAACCAGGCC	ACCCTCCGGG	CCATCACCAG	79980
	CAACGTGAGC	GCCATCCTCG	CCCGCAACGG	GGGCATCGGG	CTGTGCATGC	AGGCGTTCAA	80040
	CGACGCCAGC	CCCGGCACCG	CCAGCATCAT	GCCGGCCCTG	AAGGTCCTGG	ACTCCCTGGT	80100
25	GGCGGCGCAC	AACAAACAGA	GCACGCGCCC	CACCGGGGCG	TGCGTGTACC	TGGAACCCCTG	80160
	GCACAGCGAC	GTTTCGGGCCG	TGCTCAGAAT	GAAGGGCGTC	CTCGCCGGCG	AGGAGGCCCA	80220
	GCGCTGCGAC	AACATCTTCA	GCGCCCTCTG	GATGCCGGAC	CTGTTCTTCA	AGCGCCTGAT	80280
	CCGCCACCTC	GACGGCGAGA	AAAACGTAC	CTGGTCCCTG	TTCGACCGGG	ACACCAGCAT	80340
	GTCGCTCGCC	GACTTTCACG	GCGAGGAGTT	CGAGAAGCTG	TACGAGCACC	TCGAGGCCAT	80400
30	GGGGTTCGGC	GAAACGATCC	CCATCCAGGA	CCTGGCGTAC	GCCATCGTGC	GCAGCGCGGC	80460
	CACCACCGGA	AGCCCCTTCA	TCATGTTTAA	GGACGCGGTA	AACCGCCACT	ACATCTACGA	80520
	CACGCAAGGG	GCGGCCATTG	CCGGCTCCAA	CCTCTGCACG	GAGATCGTCC	ACCCGTCTCT	80580
	CAAACGCTCC	AGCGGGGTCT	GCAACCTGGG	CAGCGTGAAT	CTGGCCCGAT	GCGTCTCCCG	80640
	GCGGACGTTT	GATTTTGGCA	TGCTCCGCGA	CGCCGTGCAG	GCGTGCGTGC	TAATGGTTAA	80700
35	TATCATGATA	GACAGCACGC	TGCAGCCGAC	GCCCCAGTGC	GCCCGCGGCC	ACGACAACCT	80760
	GCGGTCCATG	GGCATTTGGCA	TGCAGGGCCT	GCACACGGCG	TGCCTGAAGA	TGGCCTGGA	80820
	TCTGGAGTCG	GCCGAGTTCC	GGGACCTGAA	CACACACATC	GCCGAGGTGA	TGCTGCTCGC	80880
	GGCCATGAAG	ACCAGTAACG	CGCTGTGCGT	TCGCGGGGCG	CGTCCCTTCA	GCCACTTTAA	80940
	GCGCAGCATG	TACCGGGCCG	GCCGCTTTCA	CTGGGAGCGC	TTTTTCGAACG	CCAGCCCGCG	81000
40	GTACGAGGGC	GAGTGGGAGA	TGCTACGCCA	GAGCATGATG	AAACACGGCC	TGCGCAACAG	81060
	CCAGTTCATC	GCGCTCATGC	CCACCGCCCG	CTCGGCCAG	ATCTCGGACG	TCAGCGAGGG	81120
	CTTTGCCCCC	CTGTTACCA	ACCTGTTCAG	CAAGGTGACC	AGGGACGGCG	AGACGCTGCG	81180
	CCCCAACACG	CTCTTGCTGA	AGGAACTCGA	GCGCACGTTT	GGCGGGAAGC	GGCTCCTGGA	81240

	CGCGATGGAC	GGGCTCGAGG	CCAAGCAGTG	GTCTGTGGCC	CAGGCCCTGC	CTTGCCCTGGA	81300
	CCCCGCCCAC	CCCCTCCGGC	GGTTCAAGAC	GGCCTTCGAC	TACGACCAGG	AACTGCTGAT	81360
	CGACCTGTGT	GCAGACCGCG	CCCCCTATGT	TGATCACAGC	CAATCCATGA	CTCTGTATGT	81420
	CACAGAGAAG	GCGGACGGGA	CGCTCCCCGC	CTCCACCCTG	GTCCGCCTTC	TCGTCCACGC	81480
5	ATATAAGCGC	GGCCTGAAGA	CGGGGATGTA	CTACTGCAAG	GTTCGCAAGG	CGACCAACAG	81540
	CGGGGTGTTC	GCCGGCGACG	ACAACATCGT	CTGCACAAGC	TGCGCGCTGT	AAGCAACAGC	81600
	GCTCCGATCG	GGGTCAGGCG	TCGCTCTCGG	TCCCGCATAT	CGCCATGGAT	CCCGCCGTCT	81660
	CCCCCGCGAG	CACCGACCCC	CTAGATACCC	ACGCGTCGGG	GGCCGGGGCG	GCCCCGATTC	81720
	CGGTGTGCCC	CACCCCCGAG	CGGTACTTCT	ACACCTCCCA	GTGCCCCGAC	ATCAACCACC	81780
10	TTCGCTCCCT	CAGCATCCTG	AACCGCTGGC	TGGAGACCGA	GCTCGTGTTC	GTGGGGGACG	81840
	AGGAGGACGT	CTCCAAGCTC	TCCGAGGGCG	AGCTCGGCTT	CTACCGCTTT	CTGTTTGCCT	81900
	TCCTGTGCGC	CGCGGACGAC	CTGGTGACGG	AAAACCTGGG	CGGCCTCTCC	GGCCTCTTCG	81960
	AACAGAAGGA	CATTCTTCAC	TACTACGTGG	AGCAGGAATG	CATCGAGGTC	GTCCACTCGC	82020
	GCGTCTACAA	CATCATCCAG	CTGGTGCTCT	TTACAACAA	CGACCAGGCG	CGCCGCGCCT	82080
15	ATGTGGCCCCG	CACCATCAAC	CACCCGGCCA	TTGCGTCAA	GGTGGACTGG	CTGGAGGCGC	82140
	GGGTGCGGGA	ATGCGACTCG	ATCCCGGAGA	AGTTCATCCT	CATGATCCTC	ATCGAGGGCG	82200
	TCTTTTTTGC	CGCCTCGTTC	GCCGCCATCG	CGTACCTGCG	CACCAACAAC	CTCCTGCGGG	82260
	TCACCTGCCA	GTCGAACGAC	CTCATCAGCC	GCGACGAGGC	CGTGATACG	ACAGCCTCGT	82320
	GCTACATCTA	CAACAACCTAC	CTCGGGGGCC	ACGCCAAGCC	CGAGGCGGCG	CGCGTGTACC	82380
20	GGCTGTTTTCG	GGAGGCGGTG	GATATCGAGA	TCGGGTTCAT	CCGATCCCAG	GCCCCGACGG	82440
	ACAGCTCTAT	CCTGAGTCCG	GGGGCCCTGG	CGGCCATCGA	GAATACGTG	CGATTACAGC	82500
	CGGATCGCCT	GCTGGGCCCTG	ATCCATATGC	AGCCCCTGTA	TTCCGCCCCC	GCCCCCGACG	82560
	CCAGCTTTCC	CCTCAGCCTC	ATGTCCACCG	ACAAACACAC	CAACTTCTTC	GAGTGCCGCA	82620
	GCACCTCGTA	CGCCGGGGCC	GTCGTCAACG	ATCTGTGAGG	GTCTGGGCGC	CCTTGTAGCG	82680
25	ATGTCTAACC	GAAATAAAGG	GGTCGAAACG	GACTGTGGG	TCTCCGGTGT	GATTATTACG	82740
	CAGGGGAGGG	GGGTGGCGGC	TGGGGAAAGG	GAAGGAACGC	CCGAAACCAG	AGAAAAGGAC	82800
	CAAAAGGGAA	ACGCGTCCAA	CCGATAAATC	AAGCGCCGAC	CAGAACCCCG	AGATGCATAA	82860
	TAACAAACGA	TTTATTACT	CTTATTATTA	ACAGGTCGGG	CATCGGGAGG	GGATGGGGGC	82920
	GCGCGTTTCC	TCCGTTCGCG	CTACTCGTCC	CAGAATTTAG	CCAGGACGTC	CTTGTAAC	82980
30	GCGGGCGGGG	GCGCGTGGGC	CCACAGCTGC	GCCAGAAACC	GGTCGGCGAT	GTCCGGGGCG	83040
	GTGATATGCC	GAGTCACGAT	GGAGCGCGCT	AAATCTTCGT	CGCGGAGGTC	CTGATAGATG	83100
	GGCAGTCTTT	TTAGAAGAGT	CCAGGGTCCC	CGCTCCTTGG	GGCTGATAAG	CGATATGACG	83160
	TACTTGACGT	ATCTGTGCTC	CACCAGCTCG	GCGATGGTCA	TCGGATCGGG	CAGCCAGTCC	83220
	AGGGCCTCCG	GGGCGTCGTG	GATGACGTGG	CGGCGACGTC	CGGCGACATA	GCCGCGGTGT	83280
35	TCCGCGACCC	GCTGCGCGTT	GGGGACCTGC	ACGAGCTCGG	GCGGGGTGAG	TATCTCCGAG	83340
	GAGGACGACC	GGGCGCCGTC	GCGCGGCCCA	CCGGCGACGT	CCGGGGGCTG	GAGGGGGGGG	83400
	TCTTCTTCGT	AGTCGTCCTC	GCCGCGGATC	TGTTGGGCCA	GAATTTCTGGT	CCACGAGATG	83460
	CGCGTCTCGA	GGCCGACCGG	GGCCGCGGTC	AGCGTAGGCA	TGCTCTCCAG	GGAGCGCGAG	83520
	TTGGCGCGCT	CCCCCGGGGC	CGCCCGGCGG	GCCTGGGATC	GGCTCGGGGC	GGTCCAGTGA	83580
40	CACTCGCGCA	GCACGTCCTC	GACGGACGCG	TAGGTGTTAT	TGGGGTGCAG	GTCTGTGTGG	83640
	CAGCGGACGA	ACAGCGCCAG	GAAGTGGGG	TAAGTCACT	TGAAGTACTG	CAGCAGGTCG	83700
	CGGCAGTGAA	TCGTCGGAAT	GTAGCCGGTG	CTGATGTCCA	ACACGATATC	GCAGCCCATC	83760
	AGCAGGAGAT	CGGTATCCGT	GGTATGCACG	TACGCGACCG	TGTTGGTATG	ATAGAGGTTC	83820

	GCGCAGGCGT	CGTCGGCCTC	CAGCTGACCC	GAGTTGATGT	AGGCGTACCC	CAGCGCCCCG	83880
	AGAACGCGGA	TACAGAACAG	GTGAGCCAGG	CGCAGGGCCG	GCTTCGAGGG	CGCGCCCCAG	83940
	GGGGCCGCCG	GGCCTGGGCC	GGCGGCCCGC	GTTCCCCGGT	CCCCCGGGGC	GAAGGCGTGC	84000
	CCGCGGCGGC	GCATGTTGGA	AAAAGGCGAA	ACTGGGCCCTG	GAGTCGGTGA	TGGGGGAAGG	84060
5	CGGCGGCGAG	GCGTCTACGT	CACTGGCCCTC	CTCGTCCGTG	CGGCACTGGG	CCGTCTGTGC	84120
	GGCCAGGATC	GCCTTGCCCC	CGAACACAAC	CGGCTCGGTA	CACTCGACCC	CGCGATCGGT	84180
	CACGAAGATG	GGGAACAGGG	ACTTTTGGGT	AAACACCCGT	AACATACTAC	AGAGACAGTG	84240
	TAGCGTGATT	GCCTCGCGGT	CGTAACCTTG	GTAGCGGCGC	TGATATTTAA	CCACCAGGGT	84300
	ATACATGACA	TTCCACAGGT	CCACGGCGAT	GGGGGTAAAG	TAGCCCTCCG	GGGCCCCGAG	84360
10	GCCCCGGCGC	TTCAACAGAT	GGTGAGTCTG	GGCAAACCTT	ATCATGCCAA	ACAGACCCAT	84420
	TCCGGCACGA	TTGTAGGTGC	GGATAGGTCT	CTCTACAGAG	CTGTATAGGT	GTGACGGTCC	84480
	GGGACACCCA	AGCCCCGCCG	CCCTGTGTAC	AGTGGCTGCG	GCGACGACCC	CGCTCCAACA	84540
	AGACGCTATC	CCGGGAAAGG	CACGCTCTTT	ATAATTCTTT	TTTATTTCCC	ATCTACGTGC	84600
	GGATTGGTGC	AACCGCCGGC	GCGCGCCGGT	GCAGGCCGAC	CATCTCTCTC	TTCCCCCCTT	84660
15	CCCCCTCCCC	CGAGCCCTCA	AAGAGGGTGT	GGCCTAACTA	GCGGAAGGCG	TATTTAACCA	84720
	GACTAGGGCG	GCGGGTCCGC	CGTAGTCCCT	GGCTCGGGTA	GCCACTGCTC	TGTGGCTCGG	84780
	GTCCCCCGGC	CCCCCTAACC	CCCATCCGGT	CCGCGTCATC	CGCCCCCTCC	GCCTGCGACA	84840
	CAAACGGCCG	CGCCTCCGGG	CCCGGTGACA	CGACGCGCCT	CGTCTCTGCG	GATTGTCCCG	84900
	GGAGCGTCGC	GGCATGGCTC	ATCTTCCCGG	CGGTGCGGCC	GCCGCCCCCC	TTTCGGAGGA	84960
20	CGCGATCCCG	TCGCCGCGCG	AGCGGACGGA	AGACTGGCCG	CCCTGCCAGA	TAGTGCTGCA	85020
	GGGCGCCGAG	CTGAACGGGA	TCCTGCAGGC	CTTTGCGCCG	CTTCGCACGA	GCCTTTTGA	85080
	CTCGCTCCTG	GTCGTGGGCG	ACCGAGGCAT	CCTTGACAT	AACGCGATTT	TCGGCGAGCA	85140
	GGTGTTCCTG	CCCCTCGACC	ATTTCGAGTT	CAGTCGCTAT	CGATGGGGCG	GACCCACCGC	85200
	GGCGTTCCTG	TCTCTCGTGG	ACCAGAAGCG	ATCCCTGCTG	AGCGTGTTTC	GCGCCAACCA	85260
25	GTACCCTGAC	CTGCGGCGGG	TGGAGCTGAC	GGTCACGGGC	CAGGCCCCGT	TTCGCACGCT	85320
	GGTGACGCGC	ATATGGACGA	CCGCGTCCGA	CGGAGAGGCC	GTGGAGCTTG	CCAGCGAGAC	85380
	GCTCATGAAA	CGCGAGTTGA	CGAGCTTCGC	GGTACTACTC	CCCCAGGGCG	ACCCCGACGT	85440
	CCAGCTGCGC	CTACGAAGC	CCCAGCTCAC	GAAGGTGGTG	AACGCCGTCG	GGGACGAGAC	85500
	CGCCAAACCC	ACCACGTTTC	AGCTCGGCCC	CAACGGCAAG	TTTTCCGTGT	TTAACGCGCG	85560
30	CACCTGCGTC	ACCTTTGCCG	CCCGCGAGGA	GGGCGCGTCG	TCCAGACCA	GCGCCCAGGT	85620
	CCAGATTCTG	ACCAGCGCGC	TGAAGAAGGC	GGGCCAGGCG	GCCGCCAACG	CCAAGACGGT	85680
	CTACGGGGAA	AACACACACC	GCACATTCTC	GGTGGTCGTC	GACGACTGCA	GCATGCGGGC	85740
	GGTCTTCCGG	CGGCTCCAGG	TCGGCGGGGG	GACCCCTAAG	TTCTTCCTCA	CGGCCGACGT	85800
	CCCCAGCGTG	TGTGTCACCG	CCACCGGCCC	CAACGCGGTG	TCGGCGGTGT	TTCTTTTAAA	85860
35	ACCCAGCGGG	GTCTGCCTGA	ACTGGCTCGG	CCGGACCCCG	GGTTCCTCGA	CCGGGAGCTT	85920
	GGCGTCCCG	GACTCTCGGG	CCGGCCCCGAC	CGACAGCCAG	GACTTCTCCT	CCGAGCCGGA	85980
	CGCGGGCGAC	CGCGGGCGCC	CAGAAGAAGA	AGGCCCTCGAG	GGCCAGGCC	GGGTCCCGCC	86040
	CGCGTTCCTG	GAACCGCCGG	GAACCAAGCG	GAGGCACGCC	GGGGCCGAAG	TTGTCCCGCG	86100
	GGACGACGCC	ACCAAGCGCC	CGAAGACGGG	CGTGCCCGCC	GCCCCACGC	GAGCCGAGTC	86160
40	GCCCCCCTC	TCCGCGAGAT	ACGGACCCGA	GGCGCGGAG	GGTGGTGGGG	ACGGCGGCCG	86220
	CTACGCGTGC	TACTTTTCGG	ACCTCCAGAC	CGGCGACGCG	AGCCCCAGCC	CCCTCTCCGC	86280
	CTTCCGGGGT	CCCCAAAGAC	CCCATAACGG	CTTTGGGTTG	CCCTGACGGC	GACGGGTGGT	86340
	GGCCGAACGC	TTCACCGCGC	CCGGGCACGC	GGGTGCGT	GTGTTAAAAA	AATAAATAAA	86400

	TGGGGTAGTG	TGTCCCCCCC	CTCCAACCAA	TATGGCTGTC	GTGTGTGGTT	CCGGGTGCG	86460
	CCTCCGTCCT	TTCCACCCCC	CTTCCCCCTC	CTTTTGTGTT	TTGCGTGCGC	TTATAAGAGC	86520
	GGCCCCGGGG	CCCTTCGCAG	CTTCACCGAG	AGCGCCGTCG	GGCCCCGGGT	GCGGGATGTG	86580
	TCGCGGGGAC	AGCCCCGGGG	TCGCGGGCGG	GAGCGGCGAA	CACTGCCTCG	GAGGGGATGA	86640
5	TGGGGACGAC	GGGCGCCCCC	GCCTCGCCTG	CGTGGGTGCC	ATCGCTCGGG	GGTTCGCGCA	86700
	TCTCTGGCTC	CAGGCCACCA	CGCTGGGCTT	CGTGGGGTCT	GTCTTCTGT	CGCGCGGCC	86760
	GTATGCGGAC	GCCATGTCCG	GGGCGTTCGT	GATCGGGAGC	ACCGGCCTGG	GGTTCCTCCG	86820
	CGCCCCCCCC	GCGTTCGCCC	GGCCGCGGAC	GCGTGTGTGC	GCGTGGCTGA	GGCTGGTCCG	86880
	CGGGGGAGCG	GCCGTGGCCC	TGTGGAGCCT	CGGGGAGGCC	GGCGCGCCTC	CGGGGGTTCC	86940
10	GGCCCCGGCG	ACCCAGTGCC	TGGCGCTCGG	GGCCGCCTAC	GCGGCGCTGC	TGGTGCTGGC	87000
	CGACGACGTC	CATCCCCCTT	TCCTCCTCGC	CCCGCGGCC	CTGTTTGTCT	GCACCTGGG	87060
	GGTGTGCTGC	GGCGGGCTGA	CGATAGGCGG	CAGTGC CGC	TACTGGTGGA	TCGACCCCCG	87120
	CGCCGCCGCG	GCCCTGACGG	CGGCGGTGGT	GGCGGGCCTC	GGGACAACCG	CCGCCGGGGA	87180
	CAGCTTTTCC	AAGGCCTGTC	CCCGCCACCG	CCGCTTTTGC	GTCTGTCTCC	CGGTGAGTGC	87240
15	TCCCCCGCCC	CGATACGCCC	CGGAGGACGC	CGAGCGGCCA	ACAGACCACG	GACCCCTGTT	87300
	ACCGTCGACG	CACCACCAGC	GATCTCCGCG	GGTCTGCGGC	GACGGGGCCG	CACGGCCCCG	87360
	AAACATCTGG	GTTCCCGTGG	TGACCTTTGC	GGGCGCGCTC	GCGCTGGCCG	CCTGCGCCGC	87420
	GCGAGGGTCT	GACGCGGCTC	CGTCAGGCCC	GGTCTGCGG	CTGTGGCCCC	AGGTGTTTGT	87480
	CGGGGGCCAC	GCGGCGGCGG	GCCTGACGGA	GCTGTGTGAG	ACCCTCGCGC	CCCGGGACCT	87540
20	CACGGACCCG	CTGCTGTTTG	CGTACGTCGG	ATTCCAGGTC	GTGAACCACG	GGCTGATGTT	87600
	TGTGGTCCCC	GACATCGCCG	TATACGCGAT	GCTGGGGGGC	GCCGTGTGGA	TCTCGCTGAC	87660
	GCAGGTGCTT	GGGCTCCGCG	GCCGCCTTCA	CAAGGACCCA	GACGCCGGGC	CCTGGGCGGC	87720
	CGCGACCCTG	CGGGGCCCTC	TTTTCTCCGT	CTACGCATTG	GGGTTTGCGG	CGGGGGTGCT	87780
	GGTGC GGCCG	CGGATGGCGG	CGAGCCGGCG	GTGCGGGTGA	TCGCCATTTT	AAATAAAAGG	87840
25	CACGAGTTCC	CCGAATACCA	CCGGCGTGTG	ATGATTTTCG	CCTACCGCTC	CGATCCCCCG	87900
	GGGGAGGGGG	GAAGGAAATG	GGGGCGGGGG	TGCCGTGGAC	GGGTATAAAG	GCCAGGGGGG	87960
	CAGGCGGGCC	CATCACTGTT	AGGGTGTTAG	GTTGGGAGGT	GGCACAAAAA	GCGACACACC	88020
	CGTGTGTGAG	TTGTCCGCGG	GAGGCGGTGG	TTTCCGGCAA	CCCTCCTCGC	TGCGCCGGGC	88080
	GCGCCCACCG	GTCTTTCGCG	GGGGCCGGGG	CTCTTCTGGT	CATGGCCCTT	GGACGGGTGG	88140
30	GCCTAGCCGT	GGGCCTGTGG	GGCCTGCTGT	GGGTGGGTGT	GGTCGTGGTG	CTGGCCAATG	88200
	CCTCCCCCGG	ACGCACGATA	ACGGTGGGCC	CGCGGGGGAA	CGCGAGCAAT	GCCGCCCCCT	88260
	CCGCGTCCCC	GCGGAACGCA	TCCGCCCCCC	GAACCACACC	CACGCCCCCC	CAACCCCGCA	88320
	AGGCGACGAA	AAGTAAGGCC	TCCACCGCCA	AACCGGCCCC	GCCCCCAAG	ACCGGGCCCC	88380
	CGAAGACATC	CTCGGAGCCC	GTGCGATGCA	ACCGCCACGA	CCCGCTGGCC	CGGTACGGCT	88440
35	CGCGGGTGCA	AATCCGATGC	CGGTTTCCCA	ACTCCACCCG	CACGGAGTCC	CGCCTCCAGA	88500
	TCTGGCGTTA	TGCCACGGCG	ACGGACGCGG	AGATCGGAAC	GGCGCCTAGC	TTAGAGGAGG	88560
	TGATGGTAAA	CGTGTGCGCC	CCGCCCCGGG	GCCAACTGGT	GTATGACAGC	GCCCCCAACC	88620
	GAACGGACCC	GCACGTGATC	TGGGCGGAGG	GCGCCGGCCC	GGGCGCCAGC	CCGCGGCTGT	88680
	ACTCGGTCTG	CGGGCCGCTG	GGTCGGCAGC	GGCTCATCAT	CGAAGAGCTG	ACCCTGGAGA	88740
40	CCCAGGGCAT	GTACTACTGG	GTGTGGGGCC	GGACGGACCG	CCCGTCCGCG	TACGGGACCT	88800
	GGGTGCGCGT	TCGCGTGTTC	CGCCCTCCGT	CGCTGACCAT	CCACCCCCAC	GCGGTGCTGG	88860
	AGGGCCAGCC	GTTTAAGGCG	ACGTGCACGG	CCGCCACCTA	CTACCCGGGC	AACCGCGCGG	88920
	AGTTCGTCTG	GTTTCGAGGAC	GGTCGCGGGG	TATTCGATCC	GGCCAGATA	CACACGCAGA	88980

	CGCAGGAGAA	CCCCGACGGC	TTTTCACCGC	TCTCCACCGT	GACCTCCGCG	GCCGTCGGCG	89040
	GCCAGGGCCC	CCCGCGCACC	TTACCTGCC	AGCTGACGTG	GCACCGCGAC	TCCGTGTCGT	89100
	TCTCTCGGCG	CAACGCCAGC	GGCACGGCAT	CGGTGCTGCC	GCGGCCAACC	ATTACCATGG	89160
	AGTTTACGGG	CGACCATGCG	GTCTGCACGG	CCGGCTGTGT	GCCCCAGGGG	GTGACGTTTG	89220
5	CCTGGTTCCT	GGGGGACGAC	TCCTCGCCCG	CGGAGAAGGT	GGCCGTCCGCG	TCCCAGACAT	89280
	CGTGCGGGCG	CCCCGGCACC	GCCACGATCC	GCTCCACCCT	GCCGGTCTCG	TACGAGCAGA	89340
	CCGAGTACAT	CTGCCGGCTG	GCGGGATACC	CGGACGGAAT	TCCGGTCCTA	GAGCACCACG	89400
	GCAGCCACCA	GCCCCCGCCG	CGGGACCCCA	CCGAGCGGCA	GGTGATCCGG	GCGGTGGAGG	89460
	GGGCGGGGAT	CGGAGTGCT	GTCTTGTCG	CGGTGGTTCT	GGCCGGGACC	GCGGTAGTGT	89520
10	ACCTCACCCA	CGCCTCCTCG	GTGCGCTATC	GTGCGCTGCG	GTAACCTCCG	GGCCGGGCCC	89580
	GGCCGCCCGT	TGTCTTCTTT	TCCACCCCTT	CCGTCCCCCG	TACCCACCAC	ACCCACCCCC	89640
	ACCCCCCGC	CGTCCCCCGG	GCGTTATAAG	CCGCCGCACT	CGCTTTTCCC	ACCGGAAAT	89700
	CCTCGGCCCG	ATCCGAACGG	CGCACGCCGC	GTGGGCTCCA	AACGCCTCCG	GAAGAGAGCG	89760
	CCCCGCCCGG	ATATTCAAGC	CCGCGGTGGT	GCTATGGCTT	TCCGTGCTTC	GGGACCCGCC	89820
15	TACCAGCCCC	TCGCCCCCGC	GGCCTCCCCG	GCGCGGGCTC	GTGTTCGCGC	CGTGCCCTGG	89880
	ATCGGCGTCG	GAGCGATCGT	CGGGGCCCTT	GCGCTCGTCG	CCGCGTTGGT	TCTCGTACCC	89940
	CCTCGGTCTT	CGTGGGGACT	CTCGCCGTGC	GACAGCGGCT	GGCAGGAATT	CAACGCGGGA	90000
	TGCGTCGCGT	GGGACCCAC	CCCCGTCGAG	CACGAGCAGG	CGGTCGGCGG	CTGCAGCGCG	90060
	CCGGCCACCC	TTATCCCCCG	TGCGGCCGCC	AAGCACCTGG	CCGCTCTGAC	ACGCGTCCAG	90120
20	GCGGAGAGAT	CGTCGGGTTA	CTGGTGGGTG	AACGGAGACG	GCATCCGGAC	CTGTCTGAGA	90180
	CTCGTCGACA	GCGTCAGTGG	CATCGACGAG	TTTTGCGAGG	AGCTCGCGAT	CCGCATATGC	90240
	TACTACCCAC	GAAGCCCCCG	CGGGTTTGTC	CGCTTCGTAA	CTTCGATACG	TAACGCCCTG	90300
	GGGTGCGCGT	GAGGCGCGCG	TCCGACGGTC	CCGCTTCTCG	CCTCTCTTCT	TCCCCCTCCC	90360
	CACCCACCC	ACCGACCAAC	GACGGCGTTT	GGCCAATACC	CTCCTTTTTT	CTTTTCTCT	90420
25	TCCCCCCCCA	AAAAAAAAAA	CAATAAACAG	CTAATTGCGT	ACGACAAACC	ATGCGGAACT	90480
	CGCTGTTTTT	TTTTCTCTGT	TTGTTACTTT	TTATTGAAAA	CAGACATACG	GGGAAAGGGG	90540
	CCGGAACCG	AGACGGTGGG	GCCGGCGGTC	GCATTTTTTT	AATGGCTCTG	GTGTGCGCCG	90600
	CGTTTTGAGCT	TCGTCAACAG	GGCGCTGAGG	GCGGCGACGT	TTGTGCGGCC	GTCGTTGGCC	90660
	AGCGCGTTGG	TCCGGGGGCG	GGCGGGCATG	GGCGACAGGC	TTAGTCCCGG	GTCGGGGGCG	90720
30	CGTGTGGCCC	CCGGAGGGGA	GAAGAGGGCA	GACCCGCCCC	AGTCGTACAG	GGGATTTTCC	90780
	GCCTCGATGT	ACGGGGAGTC	CGGGGCGTCT	CCCGCGGGG	CCGCCCCGCC	GGCGTCTTGC	90840
	CGGCGAAGGC	AGATGTTTTT	GTATACCCGA	ACCCAGGGGA	TCTCCTCGTA	GACGCGCCCC	90900
	CCATCCTCGC	TCACCGACTC	GTAAATGGAA	TC'TGCGTCTT	CGGAGGGGGC	GCGGGGGGCG	90960
	TGGCTTTTCG	CCGGCCAGGC	GGCGGCGGTG	GTGTGCGCGG	CGGGGGTGGC	GCCAAGCCCC	91020
35	ACGCCCCGCG	GCATGGCGGG	GTCATCGTCG	GGCAGCAGAT	ACGTGTTTTT	CATCTGGTCC	91080
	GGTTCGGCCT	CCGCGTCTGG	CCCCCAGGTC	CGCACCGCGT	CGTAAACCCC	GGCGGCCTCG	91140
	CGCTGAGCCG	CGAGCGGGCG	CGCCGCGGCT	GCCGGCCGCT	GCTCGGGGGG	CGCGGGGTTC	91200
	CGGGGCGGGA	GGCGCGGGGG	CGCCCCGGCC	ATATGCGTGT	AATACGTGGC	CGGCCGCGCG	91260
	GCGCAGGGCT	CGGGACCCCG	GTCGGCCGCG	TCGACGTGCG	GGGGCTCGGG	GAGGTCTCTG	91320
40	CGGTGCGGCC	TGAACCTCCG	AGGGGCCGCG	GGGGTCGAGT	GGGGGCGAGC	CCGGGGGAGC	91380
	GGCGGGGGTG	CGTTATCGCG	CCGGGTCCGT	TGTATCTTGT	CCCGGCAGCT	CCCGCCGACC	91440
	GCGCCGCGGC	CCCCCGGTGG	GCCGGACGCC	GCGAGGCGCA	GGATGGACTC	GTAGTGGGGC	91500
	GACGGGGTTC	CGCTCCGAAG	CAGGTCCGGG	GCCAGGGCGG	CCCCGAACCA	GGACTTGATG	91560



	CTGAGTTCCA	TCCGGGCCCA	GCTCGGGGCG	GTCATCGTGG	GGAACAGGGG	GGCGGCGGTC	91620
	CTGCAGAAGC	GCTCCTGGCT	GTCCACCGCC	GCCCGTAGGT	ACTCGTTGTT	CAGGCTGTCTG	91680
	GAGGCCCAGA	CGACATACCC	GGTAAGCGTC	GCGTTAATTA	TATACTGGGC	GTGGTGGTGG	91740
	ACTATGGATA	GAACCTCGAC	GGTCGAGACG	ATGGCGTCCA	CGATCCCGTA	CGTGCCGCGC	91800
5	CTGCGCTTGC	CGGTCTCCCA	CAGGTGGGCC	AGGCGCGTCA	GGTGGCCCAG	GACGTCGCTG	91860
	ACCGCCGCCC	GCAGGGCCAT	GCACTGCATC	GAGCCCGTGG	TGCCGCTGGG	CCGCGGTCC	91920
	AGGTGGCGCG	CAAACGTCCTC	CGCGGGCGCC	TCCAGACTCC	CGCTGAGCGC	CACGAACCGG	91980
	CGATCGGCGG	GGCCCAGGCG	GCGACACACG	TACTTGTCCG	CCGTCCACAG	CATCCACGAG	92040
	GCCCAATGGT	ACAACACGGA	GACGTAGGCC	AGGAACTCGC	TCAGCCGCAG	TGCGGTGTCC	92100
10	GTGCTCGGCC	GGCTCGGGTC	TGCGGGGCGC	ATAAAGAACA	TGTACTGCTG	GAGCCTGTGG	92160
	GCCGCGTCGC	GCAACCCCGC	CACCGCGGCG	GCGTACTTGG	CCGCGGCGGC	CCCCTCTTG	92220
	AACGGGGCGC	GCACCAGCAG	CTTCGGGAGC	AGGGTGGGCC	GCAGCAGCAC	GTGCAGGCTG	92280
	GGGTGCGAGT	CGCCCGCCGG	GTCGTCGGGG	ATGTCCAGGC	CGCTGGGCAC	GACCGTCTGG	92340
	AGGTACTTCC	AGTACTGCGC	TAGGATGGCG	CGGCTCAGCT	GGCCGCCCCG	CAGCTCCACC	92400
15	TCGCCGAGCG	CCTGCTTGGC	GGCCGACGCG	TAGTGCCGGA	TGTAGTCGTA	GTGCGGGTCG	92460
	CTGGCGAGCC	CGTCTACGAT	CAGGCTCTCG	GGGACGGTGT	TATGGTGCCG	CGCCGCCAGC	92520
	CGGACGCTGC	GATCGGCGCC	GGTCAGAAAC	GCCGGCTGCA	GGTCGTGCGC	GCGCTGCCGC	92580
	AGGACGCCCC	CGGCCGCGCT	GAGGAGCCCC	TCCGGGGTGG	GGAGCAGACA	CCCGGCGAAG	92640
	ATGCGCCGCT	CGGGGACGCC	CGCGTTGGCG	CCGCGGATGA	GGTTGGCCGG	CGTCAGGCAC	92700
20	CGCGCCAGCC	GCAGGGAGCT	CGCGCCGCGC	GCCCGGCGTT	GCATGGCGGA	GACCGTTCGG	92760
	TCGGGGGCCC	GCCGGTCGGA	GGTATGCCGC	GTCCCGGGAT	ATAGGGTTGC	TTTTTATGGG	92820
	GAGGCGCCTA	TGGGCGTGGC	GGGCCGCCCA	GCCCGGTGCG	GCGCCTCCCG	GACACGTGCG	92880
	CCCGGAGGGC	GGCGGTCTCC	TCGTGCCCCA	TGAGCAGTTT	CCGAAACTGC	GCCATGATGT	92940
	CCACGACGCG	GACCCGCGGC	CCCAGCACGG	ACTCGCTATT	CAGGGGGGCG	GGGGGGAAGG	93000
25	CCGCCAGGTC	TTCGAGCAGG	AAGGCGGGGT	CTGCCGTCCC	GCTCACGGGC	GCCCGGGGCG	93060
	CCGAGGACGC	GGGGCGAAGG	TCCACGTGTT	CCGCGGCGGC	GCGCACGTCC	GCCCAAAATT	93120
	TGGCGGGGGT	GGTCCGCGCG	TACAGGGGCT	GGGTGCGCGG	GAGGACGCAC	GCGTAGCGCA	93180
	GGGGGGTGTA	CGTGCCCAAC	TCGGGGGCGG	TCGACCCGCC	GTCAAACGCG	GCCAGGGCCA	93240
	CGCACGCGAC	CACCGTGTCG	GCCAGGCCCA	GCAGCCGCTG	CAGGATGAGC	CCCGTCGCCA	93300
30	GCACGGCGCG	CGCGGCCGCG	GCGTGGTCCC	TGCGCCGGCG	CGCGTCCCGG	CAGGCCAGGG	93360
	CGTATTTTCA	GGTAACGGTC	GCCAGGGCCG	TGTGCAGCGC	GTACACGGCC	GCGCCCAGCA	93420
	CGGCGTTTCA	CCCGCTGGTG	GCGAGCAGGC	GGCGCGCCGC	GGTGTGCCCC	AGCGCCTCGT	93480
	GCTCGGCCGC	CACGACCCCG	GGGCTACCCA	GGGGCAGGGC	GCGAAACAGC	GCCTCTGCT	93540
	CCACGTCCAC	AAACGCGGGG	TGGGCGGAGT	GCGGGTGAGC	GCGCGCCCCC	ACGACCACCG	93600
35	AGAGCCACTG	GACCGTCTGC	TCCGCCAGGA	CCGCCAGCAC	GTCCAGGACG	CGCCCCGCAA	93660
	ACGCGGCCCTC	CCGCGGGAGC	ACGCATTTGA	CGCGCGCCGG	GTTGAAGCGG	GCGAGCAGAG	93720
	CCCCGGTGGC	GATGTACGTC	ATGCGCCCCG	CGTAGCGGGC	GGCCACGCGA	CAGTCGCGCC	93780
	CCAGGAGCGC	GCGCACCCCG	GGCCAGTACA	GCAGGGACCC	CAGCGAACTG	CGAAAGACCG	93840
	CGGCGTCGGG	GCCGGGGTGG	GGGGGCGCGG	CCCCTCCCGC	GCTGAGCAGC	GGCACGGCGG	93900
40	CGGCCCCCAC	GGGCCGCAAC	GCCGTGAGGC	TCGCGAACTG	CCGTGCGAGC	TCGGCCGCCC	93960
	TGTGTCGAG	CTCCGAGCCG	CGCGCCCTCC	GTGTGCAGGC	GCGTCCCGCA	GACCCACCCG	94020
	TTGATCGCCA	CCCGCACGAT	GGCGTCCACC	AGAAAGCCCC	TCGCGCGGGA	GGGGCTGGTT	94080
	TTTGCCCGCC	GATCCGTCAG	GTCGAGGATC	GCGTCGCCCC	TGACGTACCA	GGCCAGCGCC	94140

	TCGCCCTGCT	GCAGCGTCTG	GCGGAAAAAC	ACCTTTGGGT	CGGCCGGGGA	GGCAAAGTGC	94200
	ATGACCCCCA	CGCGCGACAG	CCCGAACGCG	CTATCCGGAC	ACGGGTAGAA	CCCGGCCGGA	94260
	TGTCCCAGGG	CCAGGGCCGA	GCGCACGGAC	TCGTCCCACG	CGGCGACTCG	GGGGGTACAG	94320
	CGGTCCAGGG	GGAATGCCGC	CTGCAGCTCC	GGGCCCCACA	CGCGGCCCGC	GAGAATCTCG	94380
5	ACCGTCGCGG	GAGGCCGCGC	CCCGGCGCCG	TCATCGTGCG	CGACGGCGGC	GGGGTAGTCG	94440
	TCCTCTCTGT	AGCTGAGCTC	GTCAGGAAC	AGCGGCGAGG	GCACCACCCG	CGAACCGCCC	94500
	ACCCGCCCCA	AAACGTGCGG	TGGGTCCATC	GGGCCCAGGT	AGCCTCCCCG	CGGGGCCCGC	94560
	GTGATGGCGC	TGTCCC GGCG	TCCGCGAACG	GA CTGGCTCC	TGGCCGTAAC	GGACCTGGGG	94620
	CGCGGAAAGG	ACGCCCGGCG	GGGGGGCGCC	GCCGCCGGGG	CCTCGGACGC	GCGTCGGGAC	94680
10	CCGGGGTGAC	CGCGGGCCTC	CCGGCGACGG	CGCGGGGGCG	GCTCTTCGCT	CGCCATCTCC	94740
	CCCGCGGCTT	CGACCTCGCT	GTCGTCTGTC	ACGTTAACAA	CCGCCCGCAG	GTACCCCAT	94800
	AACCCGACTC	CACCGCCCTC	GGGCTCTGTC	TCCACGGGCG	AGTCGGCGCG	ATGCGCGGAC	94860
	GGGGCATGGG	ACCGGGTGGA	GCGCGCCTC	CGGCGTACGG	CATGCCCGCG	CACGGACATG	94920
	GTGGCCGGAG	GCCCCATTTT	TTACACACGC	CCTCCCCGCA	GACGGACGAG	GAAAGGGGTG	94980
15	GTGCGAGGGG	GGAGGCCCAA	ACGGGGAGGT	GGGGGGTAGG	GGGCGGTCCC	AGGGAGCGGG	95040
	GGGTAGGAAC	CGGCACGACG	GGAACAGAGA	AACGCGACCG	CTCCAACAAG	GGTGGGGGGT	95100
	GGGCCTCGTC	CCCACGCAGA	CCCGCGGGCA	AATGCGAGAA	CGGGACCCCG	GCGCCTGCCT	95160
	TTATACGCGG	ACCCACGAC	CACGAGCCGT	TCTGTGACGC	GAATCTACAC	GACCGCGGGC	95220
	TCGTAGGCGC	GACTAACGCC	CAACCCAACG	GCACACACCC	CCCACCCCGC	GCGTAACCCC	95280
20	ATTTCTTTCA	TGGTCCCGTA	ATAAACAGCC	AACGCACGCC	GCGTATGATG	AGTTGCTTGC	95340
	CAATGTTTAT	TGCTGTGGTT	GCGAACCCCT	TATCGCGATA	CAGACGGAAG	TGAGGCGGGG	95400
	CGGTGTGGG	GGGGGGGGGC	GCGCCGCCCG	GTCGCACATC	CTACCCCCCA	AAGTCGTCAA	95460
	TGCCCATGGC	ATCGGTAAAC	ATCTGTTCAA	ACTCAAAATC	GTCCACGTCC	AAAGCCCCAT	95520
	ACGAGACGGG	GTCGTGGGTC	ATTCCCGGGG	AGGGGGACTC	CACGTCCCCC	AGCATCTCCA	95580
25	AGTCGAAGTC	GTCCAGGGCG	TCGGCGGGCG	TCATATCCAC	CTCTCGCCG	TCCAGGCGGA	95640
	GTTCTGCTCC	CAGGCTGACG	TCGGTAATGG	GGGCGGTGGT	GGACAGTCTG	CGGGGGCGTT	95700
	GTCCCGCGGA	GAGAAACGAC	ATGCGCGGCG	CCACCAGCCC	GGCCTCCGCG	GGAGCGTCAT	95760
	CGTCGTCCGG	GAGGTCGAGC	AGGCCCTCGA	TTGTGATCC	GTAATTATTT	CTGGTCCGCC	95820
	CGCGGCTATA	CGCGTGCTCC	CGCATGACGG	ACTCGCCCTC	CGAGGTCGCA	ACGCTGGAGT	95880
30	ACGAGTCCAA	CTTGCCCCGG	ATCAGCAGCA	TAAAGTACCC	AGAGGAGCGG	GCCTGGTTGC	95940
	CCTGCAGGAC	GGGCGGGGTC	GTGAGGGGCG	CCCCGGGTTC	CTCCGCCGCC	GCACTTCGCA	96000
	CCAGCGGGAG	GTTCAAGTGC	TCGCGAATGT	GGTTTAGCTC	CCGCAGTCGC	CGGGCCTCCA	96060
	CGGGAACTCC	CCGCACGGTG	AGCGATCCGT	TGATAAACAT	CAGGGGCTGA	AACAGACACG	96120
	CCAAC TGGCG	CCAGCTCTCC	AGGTGCGAGC	AGAGGCCGTC	GAACAGATCG	GGCCGCATCA	96180
35	TCTGCTCGGC	GTACGCGGCC	CATAGGATCT	CGCGGCTCAG	AAAGAGGTAT	AGATGCAGAA	96240
	ACAGGACCGG	CGCCAGGCGC	GCGGTCTCGC	GGTAGTACCT	GTCCGCGATC	GTGGTGCGCA	96300
	GCATCTCCCG	CAGGTCGCGG	TGCGGGCCCC	GCATGTGTGC	CTGGCGGTGT	AGCTGCCGAA	96360
	CGCTGGCGCG	CAGGTACCGG	TACAGGGCCG	AGCAAAAATT	TGCCAACACG	GTCCGGTAGC	96420
	TCTCTCTCCG	CGCCCGCAGC	TCACCGCGGA	AAAAC TGC	CATGGCCTCG	TAGTACGAAG	96480
40	GCAGCTCGTC	GCGGGTGGCG	GGCAGGGTGG	GGAACGCCAC	GTCGCCGTGG	GCGCGAATGT	96540
	CGATCGGGGA	GCGCTCGGGG	ACGTGCGCAT	CCCCCAGTC	GATCACGTCG	CTGGGCAGCG	96600
	TCGACAGAAA	CTTGCACTCC	CGGTACATGT	CGGCGTTGGT	CGGGAACCCA	GAGAACAGGT	96660
	CCTCGTTCCA	GGTATCTAGC	ATGGTACACA	GCGCGGGACC	CGCGCTGAAG	CCCAGATCGT	96720

	CGAGGAGACG	GTAAACAGG	GCCGCGGGG	GGACGGGCAT	GGGCGGCGAG	GGCATCAGCT	96780
	GGGCCTGACT	CAGCCGACCG	GTGGCGTACA	GCGGAGGGGC	GGCTGGGGTG	TTCTTGGGAC	96840
	CCCCGGCTGG	CCTGGGGGGC	GGTGGCGAAA	CCCCGTCCGC	GTCCGCAAAC	AGATCGTCGA	96900
	CCAACAGGTC	CATGGGGGCG	GTTGGGTCCG	GGAATAACGA	TCTCGAGAGG	CGAATGAGAC	96960
5	GTGCCCAGAC	GCCCGGCGGC	GGAGAGGGGG	GGAGGGATCC	GGGACCCGCG	ACAGAAAAAG	97020
	GCCGGGGCCC	TCGCGAAGGG	AATCGCCGGG	GGTGCCGTGC	GTCCCCGAGG	ACTGACATCT	97080
	CGCGTCCACC	ACCCCGCATT	TAAGTATCAC	CCCAGTGCCG	CCCCAAACCT	CGTGAATTCC	97140
	CCACCGCTCC	GGGCGGCCCG	TCCCCCGCGC	TCGGAAGGGA	GGCGTGTGCG	CCCTCCCGCC	97200
	CCTCCCGCCC	CTCCCGCCCC	TCCCGCCCCCT	CCCGCCCCCTC	CCGCCCCCTCC	CGCCCCCTCCC	97260
10	GCCCCCTCGCC	ACAAACGCGT	GCTGACAGCG	AAGTGGTTAA	ATCGACCGTG	ATGCTTTATT	97320
	GTCTGTCTGTC	TGAACGCGGC	CGGGGTCTGCT	ACTCGAGGGG	GCGGCGGGGA	CGGGAAGCCG	97380
	AGCGGGCGGG	GGCCCGTGCG	GTCGCGGCGG	CACGCCCCGC	GGGGCGGGCC	CGGGCGGGCG	97440
	CGGTCTGCGTC	GACGTCTCTG	GCCGCGTCCG	GATTACACAA	CTCGTTCGCG	CGCTGCAGGA	97500
	GGTTCCTTGCC	CTCGCAGACC	GTCACGCGAA	TGGTGGTGAG	GTGAGGAGC	TCGTTGAGGT	97560
15	CTTCGTCGGT	GTGCGGCCGC	GACATGTCCC	ACAGCTGTAC	CGCCGCCAGC	CGGGCGTGCG	97620
	TGGCCGCCAG	GCGCCCGACC	GCGGCGCAGA	AGACGCGCTT	GTTGAACCCG	GCCACCCGGG	97680
	GGGTCCACGG	CGCCGTGGGG	CTCGGTGGGG	CGGTGCTGAA	GTGCAGCTTC	TTGGCCAGTC	97740
	CCTGGGCGGG	TGTCTTGTTT	CTTCCCGAGG	CCGTGGGAGC	GGGGGCGTCT	AGGAGCACGG	97800
	CGGAGTCGGC	CTGGGCGGGT	CGCCTGCCGC	GGGAGGGTTC	GGTCGCGGGG	GTCGCGGGGG	97860
20	CCTTAGCGCG	CCCGCGCGTC	ATTTTGGGGG	TCCGCGCGGG	AGGGGCGTGC	GAGCGCCCCG	97920
	CGGCGCCAC	GGGGCCCCCG	GGGGGTGGAG	GAGCGCGCGC	GGGGCCGGGG	CCGTGAGAGC	97980
	CCGCGACGGA	CGCCGAACGA	CGCGGTCCGC	CGGTATCCCG	GGACTCGTCG	TTGTCTTCGG	98040
	ACGACGACGA	GTCCCGGTAG	AGGGCATAAC	CAGCCTCGTC	ATAATGGAGA	AAGCGAACCT	98100
	CGCCCCTTGG	GCGCGCGCGC	ATCGGGCCAG	CGCCGCGGCG	GAAGTCGTCG	CGCGGACTCT	98160
25	CTGGATCCGC	CGGGGAGACC	GGGCCATAGT	ACAGCTCCTC	GTGGGTCCCG	CGCGGCGCTT	98220
	CCCGCGGACA	CGACTTGACG	GAGCGGCGAG	AGGTATGGT	CTATCGGAGA	CACCGGGGAC	98280
	GCCCCGTGCG	ATCACAGGGA	AGGCGTCCGC	GAAGCAGGCA	GAGAGCGTCG	GAAGGCGGCG	98340
	AGGGAGGGAA	AGAGGGAGAC	CGGCGGGGTA	CGGGAGAGCA	GCGAGGGCCT	GCGTAACCCA	98400
	CGGGGGCCGC	GGGAGTGGCT	CCCTGCGGGT	TGCGGGGGAG	AGTTTATAGG	AAGTGGATAT	98460
30	AACCGCAGGC	GACGGGACTA	ACCAATCCCC	GGGGGGGCAA	CGGACAGACA	CGCCCCGAAC	98520
	CGGCCCGACT	TCCGCGAGGA	AGCAAAGGCC	GGGGGCCGCC	CAACGACACG	CCCACCCCTT	98580
	CCCAACAGGG	CGGGCTCAGG	CTGACCCGGC	GGCCAGTGCC	CGCTGGCATA	TCTGATACAC	98640
	GTGCGCGATC	ATACATACGC	CCATCGAGGT	CATGCCTAGA	TAAAAGGGCA	CCAGGACCCC	98700
	CGGGACGGAC	ACCACACCGG	CGCTGTGCGC	CCGGCATTGC	GCGTCCCCGA	TAACGCCGCG	98760
35	TGCGCCTGCC	GCGTTCGGCG	GCTCCCCGGG	CACGCCCGCG	ACGAGCGCGA	CGAACAACAG	98820
	CACCACCCAG	CGGCCAGTC	TTGCGGGTTT	CCCCGTCATC	GCGGCGATGA	GTCAGTGGGG	98880
	GCCCAGGGCG	ATCCTTGTC	AGACGGACAG	CACCAACCGG	AATGCCGATG	GGGACTGGCA	98940
	AGCGGCCGTA	GCTATTGCGG	GGGGCGGAGT	CGTTCAACTG	AACATGGTCA	ACAAACGCGC	99000
	CGTGGATTTT	ACCCCGGCAG	AATGCGGGGA	CTCCGAATGG	GCCGTGGGCC	GCGTCTCTCT	99060
40	GGGCCTGCGA	ATGGCAATGC	CGCGTGACTT	CTGCGCGATT	ATTACGCCC	CCGCGGTATC	99120
	CGGCCCGGG	CCCCACGTGA	TGCTCGGTCT	CGTCGACTCG	GGCTACCGCG	GAACCGTCCT	99180
	GGCCGTGGTC	GTAGCCCCGA	ACGGGACGCG	CGGGTTTGCC	CCCGGGGCC	TCCGGGTGCA	99240
	CGTGACGTTT	CTGGACATCC	GGGCCACCCC	CCCGACCCCTC	ACCGAGCCGA	GCTCCCTGCA	99300

	CCGGTTTCCG	CAGTTGGCGC	CGTCCCCGCT	GGCAGGGTTA	CGAGAAGATC	CTTGGTTGGA	99360
	CGGGGCGCTC	GCGACCGCCG	GGGGGGCGGT	GGCCCTGCCG	GCCAGACGGC	GCGGGGGATC	99420
	GCTGGTCTAC	GCGGGCGAGC	TAACGCAGGT	GACCACCGAG	CACGGCGACT	GCGTGACCA	99480
	GGCGCCCGCC	TTTCTGCCAA	AGCGCGAGGA	GGACGCAGGC	TTTGACATTC	TCATCCACCG	99540
5	AGCCGTGACC	GTCCCGGCCA	ACGGCGCCAC	GGTCATACAG	CCGTCCCTCC	GCGTATTGCG	99600
	CGCGGCCGAC	GGACCAGAGG	CCTGCTATGT	GCTGGGGCGG	TCGTGCTCA	ATGCCAGGGG	99660
	CCTCCTGGTC	ATGCCTACGC	GCTGGCCCTC	CGGGCACGCC	TGTGCGTTTG	TTGTATGTAA	99720
	CCTGACCGGA	GTCCCGGTGA	CCCTACAAGC	CGGGTCCAAG	GTGCGCCAGC	TGCTCGTCGC	99780
	GGGGACCCAC	GCCCTCCCTT	GGATCCCCCC	CGACAACATC	CACGAGGACG	GCGCATTCCT	99840
10	GGCCTACCCC	AGAGGGGTTC	CGGACGCGAC	CGCCACCCCC	CGAGACCCGC	CGATTTTGGT	99900
	GTTTACGAAC	GAGTTTGACG	CGGACGCCCC	CCCAAGCAAG	CGGGGGGCGG	GGGGGTTTGG	99960
	CTCCACTGGC	ATCTAGACCG	CGCCTCGCGT	CGGGCCAGAT	GGGGCCCCCG	TCAATAAAGA	100020
	GCTCTGTTTC	GCATATGCCC	TGGTGTGGGC	GGTTTTTTTT	TTGTTGTCTG	TCTGCCCGGC	100080
	ACTCGGTTGT	CCGTTCTGTC	GTCGCTATCA	CATACGCACA	AACACACGGG	TAGAGTGGAA	100140
15	CCGAAACCGG	TCGACGTTTA	TTCACCACAC	AGAAACACAA	GCTAAGCGAG	AAGGAGGGGG	100200
	GCCTCGGTCG	ACGAGGCCCTG	GCGTTTGGGG	GCGGACGTGC	GATGACGTGG	GTCCGGTGTA	100260
	GGGTCCGCGG	GGGGCACGGG	CCCGGGGCGA	ACGGGGGATC	TGTCGCCGGC	GTGGGTGACT	100320
	GGGACCGACG	CAACCTCCGG	GGCTTGTGCC	CTCGTAGGCC	CGGGGGGGGC	CTCGGTGCT	100380
	CCGAGCCCCG	CGGTGCGGGT	CCCTCCGGCC	AGAGCCGAGG	TGGAGAGACC	AAGGGCCCGC	100440
20	TCCGCGATCG	CCACGTCCTC	CATGACCACG	TCGCTCTCGG	CCATGCTCCG	AATGGCCTGG	100500
	GAGACGAGCA	CGTCCGCCGA	CTTGTCCGCG	GCCCCACCG	ACATGTACAT	CTGCAGGATG	100560
	GTGGCCATGC	ACGTGTCCGC	CAGGCGGCGC	ATCTTGTCCC	GATGCGCCGC	AACGGCCCCG	100620
	TCGATGGTGG	AGCCCTCGAG	TCCCGGGTGG	TGGCGCGCCA	GCCTCTCGAG	GTTGACCATG	100680
	CAGGCGTGGT	ATGTGCGGGC	CAGGGCGCGC	GCCTTCACGA	GGCGCCGGGT	GTCGTCCAGC	100740
25	GACTCTAGGG	CGTCGTCGAG	CGTGATGGGG	GCGGGCAAAA	GCGCATTGAC	CACCGCCAGG	100800
	GCCTCCTGCA	GCCGCGGGTC	CNNNTCCGAG	GGCGGAGCCG	CGGCCGAAT	CATCTCATAT	100860
	TGTTGTTTCT	CGGGGCGCGT	TCCCCAACCG	CACAGCACCC	CGAGCAGGGA	CGCCATCCCG	100920
	GAACACGCGC	GCGGCTCTGC	GCCGGCTTTC	CCCCACCCCA	CCCCCTCCGG	GTTGCGAGGG	100980
	GCGATGGGGA	CGGAAGACTG	CGATCACGAA	GGGCGGTCCG	TTGCGGCTCC	CGTGGAGGTT	101040
30	ATGGCGCTGT	ATGCGACCGA	CGGGTGC GTT	ATCACCTCCT	CGCTCGCCCT	CCTCACAAAC	101100
	TGCCTGCTGG	GGGCCGAGCC	GTTGTATATA	TTCAGCTACG	ACGCGTACCG	GCCCCGATCG	101160
	CCCAATGGCC	CCACGGGCGC	GCCCACCGAA	CAGGAGAGGT	TCGAGGGGAG	CCGGGCGCTC	101220
	TACCGGGATG	CGGGGGGGCT	AAATGGCGAT	TCATTTCTGGG	TGACCTTTTG	TTTATTTGGG	101280
	ACGGAAGTGG	GCGTGACCCA	CCACCCGAAA	GGGCGCACCC	GGCCCATGTT	TGTGTGCCGC	101340
35	TTCGAGCGAG	CGGACGACGT	CGCCGTGCTC	CAAGACGCCC	TGGGCCGCGG	GACCCCATTG	101400
	CTCCCGGCCC	ACATCACAGC	AACTCTGGAC	TTGGAGGCGA	CGTTTGCGCT	CCACGCTAAC	101460
	ATCATCATGG	CTCTCACCGT	GGCCATAGTC	CACAACGCCC	CCGCCCGCAT	CGGCAGCGGC	101520
	AGCACCGCTC	CCCTGTATGA	GCCCGGCGAA	TCGATGCGCT	CGGTGCTCGG	GCGCATGTCC	101580
	CTGGGGCAGC	GCGGCCTCAC	CACGCTGTTC	GTGCACCACG	AGGCGCGCGT	GCTGGCGGCG	101640
40	TACCGCCGGG	CGTATTATGG	GAGCGCCCAA	AGCCCCTTTT	GGTTTCTGAG	CAAATTCGGC	101700
	CCGGACGAAA	AGAGCCTGGT	GCTGGCCGCT	AGGTACTACG	TACTCCAGGC	TCCGCGCTTG	101760
	GGGGGCGCCG	GAGCCACGTA	CGATCTGCAG	GCCGTGAAAG	ACATCTGCGC	GACCTACGCG	101820
	ATCCCCACAG	ACCCACGCCC	CGACACCCCT	AGTGCCGCGT	CCTTGACCTC	GTTGCGCGCC	101880

	ATCACTCGGT	TCTGTTGCAC	GAGCCAGTAC	TCCCGCGGGG	CCGCGGCCCG	TGGGTTTCCG	101940
	CTGTATGTGG	AGCGCCGCAT	CGCCGCCGAC	GTACGCGAGA	CCGCGCGGCT	GGAGAAGTTC	102000
	ATCGCCACG	ATCGCAGTTG	CCTGCGCGTG	TCCGACCGGG	AATTCATTAC	GTACATCTAC	102060
	CTGGCCCACT	TTGAGTGCTT	CAGCCCCCCG	CGCCTGGCCA	CGCATCTCCG	GGCCGTGACC	102120
5	ACCCACGACC	CCAGCCCCGC	GGCCAGCACG	GAGCAGCCCT	CGCCCCCTGG	TCGGGAGGCG	102180
	GTGGAACAGT	TCTTCCGGCA	CGTGCGCGCC	CAGCTGAACA	TCCGCGAGTA	CGTAAAGCAA	102240
	AACGTCACCC	CCAGGAAAC	CGCCCTGGCG	GGAGACGCGG	CCGCCGCCCTA	CCTGCGCGCG	102300
	CGCACGTATG	CCCCGGCGGC	CCTCACGCCC	GCCCCCGCGT	ACTGCGGGGT	CGCAGACTCG	102360
	TCCACCAAAA	TGATGGGACG	TCTGGCGGAA	GCAGAAAGGC	TCCTAGTCCC	CCACGGCTGG	102420
10	CCCCGCTTCG	CACCAACAAC	CCCCGGGGAC	GACGCGGGGG	GCGGCACTGC	CGCCCCCCAG	102480
	ACCTGCGGAA	TCGTCAAGCG	CCTCCTCAAG	CTGGCCGCCA	CGGAGCAGCA	GGGCACGACG	102540
	CCCCCGCGCA	TCGCGGCTCT	CATGCAGGAC	GCGTCGGTCC	AAACCCCCCT	GCCCCGTGTAC	102600
	AGGATTACCA	TGTCCTCCGAC	CGGCCAGGCG	TTTGCCGCGG	CGGCGCGGGA	CGACTGGGCC	102660
	CGCGTGACGC	GGGACGCGCG	CCCGCCGGAA	GCGACCGTGG	TCGCGGACGC	GCGGCGGCGG	102720
15	CCCGAGCCCG	GCGCGCTCGG	CCGGCGGCTC	ACGCGCCGCA	TTTGCGCCCG	GGGCCCCGCG	102780
	CCTCCCCCGG	GGCGGCCTGG	CCGTGCGGGG	CCAGATGTAC	GTGAACCGCA	ACGAGATCTT	102840
	CAACGCCGCG	CTGGCCGTTA	CGAACATCAT	CCTGGATCTG	GACATCGCCC	TGAAGGAGCC	102900
	CGTCCCTTTT	CCCCGGCTCC	ACGAGGCCCT	GGGTCACTTT	AGGCGCGGGG	CGCTGGCGGC	102960
	GGTTCAGCTG	TTGTTTCCCG	CGGCCCGCGT	AGACCCCGAC	GCCTATCCCT	GTTATTTTTC	103020
20	CAAAAGCGCC	TGTCGGCCCC	GCGCGCCGCC	CGTCTGTGCG	GGCGACGGGC	CCTCGGCCCG	103080
	TGGCGACGAC	GGCGACGGGG	ACTGGTTCCC	CGACGCCGGT	GGCGACGACG	GCGACGAGGA	103140
	GTGGGAGGAG	GACACGGACC	CCATGGACAC	GACCCACGGC	CCCCTCCCGG	ACGACGAGGC	103200
	CGCGTACCTC	GACCTGCTAC	ACGAACAGAT	ACCAGCGGCG	ACGCCCAGCG	AACCGGACTC	103260
	CGTCGTGTGT	TCCTGCGCCG	ACAAGATCGG	GCTGCGCGTG	TGCCTACCGG	TCCCCGCCCC	103320
25	GTACGTTGTG	CACGGCTCCC	TGACGATGCG	TGGGGTGGCG	AGGGTGATCC	AGCAGGCGGT	103380
	GCTGTTGGAC	CGCGACTTCG	TGGAGGCCGT	AGGGAGCCAC	GTAAAGAAGT	TTTTGCTGAT	103440
	CGATACGGGC	GTGTACGCCC	ACGGCCACAG	CCTGCGCTTG	CCGTATTTTC	CCAAGATCGG	103500
	CCCCGACGGC	TCCGCGTGCG	GCCGGTTATT	GCCCGTCTTC	GTGATCCCCC	CCGCGTGCGA	103560
	GGACGTTCCG	GCGTTCGTGC	CCGCGCACGC	CGACCCGCGG	CGCTTCCACT	TTCACGCCCC	103620
30	GCCCATGTTT	TCCGCGGCCC	CGCGGGAGAT	CCGCGTCTC	CACAGCCTGG	GCGGGGACTA	103680
	TGTCAGCTTT	TTCGAGAAGA	AGGCGTCGCG	CAACGCCCTG	GAGCACTTTG	GGCGACGCGA	103740
	GACCCTGACG	GAGGTTCTGG	GCCGCTACGA	TGTGCGGGCC	GACGCGGGGG	AGACCGTGGA	103800
	GGGGTTCGCG	TCAGAACTGC	TGGGGCGAAT	AGTCGCGTGC	ATCGAGGCCC	ACTTTCCCGA	103860
	GCACGCGCGG	GAATATCAGG	CCGTGTCCGT	TCGCCGGGCC	GTCATTAAGG	ACGACTGGGT	103920
35	CCTGCTGCAG	CTGATCCCCG	GCGCGGCGC	CCTGAACCAA	AGCCTCTCGT	GTCTGCGCTT	103980
	CAAGCACGGC	AGGGCAAGTC	GCGCGACGGC	CCGGACCTTT	CTCGCGCTGA	GCGTCGGGAC	104040
	CAACAACCGC	CTATGCGCGT	CCCTGTGTCA	GCAGTGCTTT	GCCACTAAAT	GCGATAACAA	104100
	CCGCTGCAC	ACGCTGTTTA	CCGTGATGCG	GGGCACGCCA	TGCTGCGGGT	CCGCTCCCTC	104160
	CAGCACCTCA	CGACCGTCAT	CTTCATAACG	GCCTACGGCC	TCGTGCTCGC	GTGGTACATC	104220
40	GTC'TTTGGTG	CCAGTCCGCT	CCACCGATGT	ATTTACGCGG	TGCGCCCCGC	CGGGGCACAC	104280
	AACGATACCG	CCCTCGTGTG	GATGAAGATA	AACCAGACGC	TGTTGTTTCT	GGGCCCCCGG	104340
	ACCGCCCCCG	CCGGCGGGGC	ATGGACCCCC	CACGCCACGC	TCTGCTACGC	CAATATCATC	104400
	GAAGTGCGGG	CCGTGTCCCT	CCCGGCCATC	CCCGCGGCCA	TGAGCCGCCG	GGTCATGAAC	104460

	GTGCACGAGG	CCGTAAACTG	CTTGGAGGCC	CTCTGGGACA	CCCAGATGCG	CCTGGTGGTC	104520
	GTCGGTTGGT	TTCTGTATCT	AGCGTTCGTC	GCCCTTCACC	AACGACGATG	CATGTTCCGC	104580
	GTCGTGAGTC	CCGCGCACAG	CATGGTGGCC	CCGGCGACCT	ATCTTTTGAA	CTACGCCGGC	104640
	CGCATAGTGT	CGAGCGTGTT	CTTGCAATAC	CCCTACACGA	AAATCACCCG	CCTCCTCTGC	104700
5	GAGCTATCCG	TTCAACGCCA	GACCTGGTG	CAGCTGTTCG	AGGCGGATCC	GGTCACCTTC	104760
	TTGTACCACC	GCCCGGCCGT	TGGCGTCATC	GTGGGCTGCG	AGCTGCTGCT	CCGCTTCGTG	104820
	GCCCTCGGTC	TCATCGTCGG	CACCGCTCTC	ATCTCCCGGG	GCGCCTGCGC	GATCACATAC	104880
	CCCCTGTTTC	TAACAATCAC	CACCTGGTGT	TTCGTGTCCA	TCATCGCCCT	GACGGAGCTG	104940
	TATTTTCATCC	TGCGGCGGGA	CTCGGCCCCC	AAAAACGCGG	AACCAGCGGC	CCCCAGGGGG	105000
10	CGCTCCAAAG	GGTGGTCGGG	CGTCTGCGGG	CGCTGCTGTT	CCATCATCCT	CTCCGGTATC	105060
	GCCGTGCGCC	TGTGCTATAT	CGCCGTCTGT	GCCGGGGTGG	TGCTTATGGC	GCTTCGCTAC	105120
	GAACAGGAGA	TTCAGCGGCG	CCTGTTTGAT	CTGTGACGTA	ACGCCTCTTC	CGTTGGAAGA	105180
	GGCGGACCCA	GTCGCCCATG	CAAATTAAAT	ACACGACCCG	CCTCGGGCCT	ACGCACCCCTC	105240
	GCACGTGCGA	TGCAAATTAA	AATCGTGCAC	AGAGCCGATC	CGGCCTCGGG	TCTGCTTGCC	105300
15	CCTCCCCCGG	TCCAGCACAG	GCAGGCTCGT	CCGACTTCCG	CATACACCCC	ACCCATACCGC	105360
	GTGCTTCCGC	ACCCCCGCCT	ACGCGTGATC	GCGAAGGCGG	ACCCAGACCT	GCCGTATGCT	105420
	AATTAAATAC	ATAAAACCCA	CCCTCGGCGT	CCGATTGGTT	TCTGGGGACG	GCGGGGGCGG	105480
	GGGCGGTGAC	GCCCGACGGG	GAGGACAAG	GAGGAGTTTC	GGAAAGCCGG	CCCCGGTCGT	105540
	GCGGGTATAA	GGGAGCCAC	CGGCCCACTG	GGCGCTGTGT	GCTGCCGTGT	GCCGACCCCG	105600
20	GTTGCGCGTC	GGTGCCGCTC	CTCGATTCCG	ACCCGGCCAC	TCTCTTCCGA	CACGCGCCCC	105660
	CTCGGAGGAC	ACCCGCCATC	CCAGCCCCCG	CGACCTACAA	CATGGCTACC	GACATTGATA	105720
	TGCTAATCGA	CCTAGGATTG	GACCTGTCCG	ACAGCGAGCT	CGAGGAGGAC	GCTCTGGAGC	105780
	GGGACGAGGA	GGGCCGCCGC	GACGACCCCG	AGTCCGACAG	CAGCGGGGAG	TGTTCCCTCGT	105840
	CGGACGAGGA	CATGGAAGAC	CCCTGCGGAG	ACGGAGGGGC	GGAGGCCATC	GACGCGGCGA	105900
25	TTCCCAAAGG	TCCCCCGGCC	CGCCCCGAGG	ACGCCGGCAC	CCCCGAAGCC	TCGACGCCTC	105960
	GCCCGGCAGC	GCGGCGGGGA	GCCGACGATC	CGCCACCCGC	GACCACCGGC	GTGTGGTTCG	106020
	GCCTCGGGAC	CAGGCGGTCT	GCTTCCCCCC	GGGAACCGCA	CGGGGGGAAG	GTGGCCCGCA	106080
	TCCAACCCCC	GTCGACCAAG	GCACCGCATC	CCCGAGGCGG	GCGGCGAGGT	CGCCGCGGGG	106140
	GCCGGGGTCG	ATACGGCCCC	GGCGGCGCCG	ACTCCACACC	AAACCCCCGC	CGGCGCGTCT	106200
30	CCAGAAACGC	CCACAACCAA	GGGGGTGCGC	ACCCCGCGTC	GGCGCGGACG	GACGGCCCCG	106260
	GCGCCACCCA	CGGCGAGGCG	CGGCGCGGAG	GGGAGCAGCT	CGACGTCTCC	GGGGGCCCCG	106320
	GGCCACGAGG	CACGCGCCAG	GCCCCCCTC	CGCTGATGGC	GCTGTCCCTG	ACCCCCCCGC	106380
	ACGCGGACGG	CCGCGCCCCG	GTCCCGGAGC	GAAAGGCGCC	CTCTGCCGAC	ACCATCGACC	106440
	CCGCCGTTTC	GGCGGTTCTG	CGATCCATAT	CCGAGCGCGC	GGCGGTCGAG	CGCATCAGCG	106500
35	AAAGCTTTGG	ACGCAGTGCC	CTGGTCATGC	AAGACCCCTT	TGGCGGGATG	CCGTTTCCCC	106560
	CCGCGAACAG	CCCCTGGGCT	CCCGTGCTGG	CCACCCAAGC	GGGGGGGTTT	GACGCCGAGA	106620
	CCCGTCGGGT	TTCTGGGAA	ACCCTGGTCG	CTCACGGCCC	GAGCCTCTAC	CGCACATTCT	106680
	CAGCCAACCC	GCGGGCCGCG	TCGACAGCCA	AGGCCATGCG	CGACTGCGTG	CTGCGCCAGG	106740
	AAAATCTCAT	CGAGGCCCTG	GCGTCCGCGG	ATGAGACGCT	GGCGTGGTGC	AAGATGTGCA	106800
40	TTCACCACAA	TCTGCCGCTC	CGCCCCCAGG	ACCCTATCAT	CGGAACGGCG	GCCGCCGTGC	106860
	TGGAACACCT	CGCCACGCGC	CTGCGCCCCC	TTCTGCACTG	CTACCTGAAG	GCCCGAGGCC	106920
	TGTGCGGGCT	GGACGACCTG	TGCTCGCGGC	GACGCCGTGC	GGACATTAAG	GATATTGCCT	106980
	CCTTTGTGTT	GGTCATCCTG	GCCCGCCTCG	CCAACCGCGT	CGAGCGCGGC	GTGTCCGAGA	107040

	TCGACTACAC	GACCGTGGGG	GTTGGGGCCG	GCGAGACGAT	GCACTTTTAC	ATCCCGGGGG	107100
	CCTGCATGGC	GGGTCTCATT	GAAATACTGG	ACACGCACCG	CCAGGAGTGT	TCCAGTCGCG	107160
	TGTGCGAGCT	GACGGCCAGT	CACACTATCG	CCCCCTTATA	TGTGCACGGC	AAATACTTCT	107220
	ACTGCAACTC	CCTATTTTAG	GCAAGAATAA	ACATATTGAC	GTCAACCCAA	GTGGTTCCGT	107280
5	GTGATGTTCT	TGGCGCGCGC	GGCGGGTGGG	GCGGAGACTC	CGGGGCGATG	CCGGCGTGCG	107340
	CGTGGGAGGA	GGGCGATGAC	CCACCGGATA	AATGTGGGGC	CCCGGCCCGG	CCCGCTTCAT	107400
	AGCGCGTCCA	GGAACACACG	GCAGACGCGT	ATTCACCGAC	CCCCCCCCCTC	GCAACATGAC	107460
	AACGACGCCC	CTCTCGAACC	TGTTTTTACG	GGCCCCGGAC	ATCACCCACG	TCGCCCCCCC	107520
	GTACTGTCTG	AATGCCACGT	GGCAGGCCGA	AAACGCCCTG	CACACGACCA	AAACGGACCC	107580
10	CGCGTGCCCTG	GCCGCGCGGA	GTTATTTAGT	CCGCGCCTCC	TGCTCGACCA	GCGGCCCCAT	107640
	CCACTGTTTT	TTCTTTGCGG	TGTACAAGGA	CTCGCAGCAC	TCCCTTCCGC	TGGTTACCGA	107700
	GCTCCGCAAC	TTCGCGGACC	TGGTCAACCA	CCCGCCCGTC	TTGCGCGAAC	TAGAGGATAA	107760
	GCGTGGGGGG	CGGCTGCGGT	GCACGGGCCC	ATTCAGCTGC	GGAACCATCA	AGGACGTCTC	107820
	CGGTGCATCC	CCCGCGGGGG	AATACACGAT	AAACGGTATC	GTGTACCACT	GTCACTGTCTG	107880
15	GTATCCGTTT	TCCAAAACCT	GCTGGCTCGG	GGCATCCGCG	GCCCTACAAC	ACCTTCGCTC	107940
	TATAAGCTCA	AGCGGCACGG	CCGCTCGCGC	GGCAGAACAG	CGACGCCACA	AAATCAAAAT	108000
	CAAAATCAAG	GTATAACCCA	CCCCCTTCCC	TCCGAGTCCG	TATGCAACCT	CATTAATAAA	108060
	GAGTGAGAAC	CAACCAAAAC	AGACGCGGTG	TGAGTTTGTG	GGTTATAGGA	ACCCGGTAAA	108120
	TACCACGCGA	CGAACCAGCG	TGTGTGTTAA	CGCGACTTTT	ATTCGTTGTA	TCGCGGGAGG	108180
20	GGGGAAGCTT	ACCGCCAAAG	GAAGGCCAAG	ATGATAACGA	CGACCACCGC	GACCACCCCA	108240
	AAAACCGCAT	GACGACACGT	CCCGCCACAC	CACCCTGGGG	CTTGGGGCGT	GTCGGAGCTC	108300
	GACGCACAGC	GGGCCGCGCG	TTGGGGCCCG	TACAGCTCTC	GCGAATTGAC	GAGCGGGGGT	108360
	CGCCACGTGC	GCGAGCTTTG	CACGCGGGGT	TGGTCCGGCC	GCCCCACGGA	CCCGCCCGGT	108420
	GGCTCGGTCTG	GACATGCGGC	CATGACCATG	GCGTAGGTGG	GGGGGCGATC	CGAGGTCGCC	108480
25	TCTGCGTAAG	TAGGGAGGCC	CGACGGGAGG	TCGCCTCCCA	CGCCAGGGTG	GGCCCCAATC	108540
	ATAGTTTCCG	GTAGAAACAG	GGGGGTCTCC	ACAAACAACC	CCCCTGGGCC	AAAGCTCCGG	108600
	CGCCGCGCCC	GTCGTTCCGC	GCGGCGCCTG	GCGCGCCGAG	CGGCCCGCCA	GGCGGCGCGG	108660
	CGCGAGCGGC	CACGCTCACA	CACCTCGCCG	TCACCGGAAG	AAGCCGGTGA	AACAAGCCCA	108720
	ACCGGCGACG	TCCCTGCAGA	GTACGGTGGA	GGCGAGTCCG	TGGGGGTGTC	GATATCAATA	108780
30	ACGACAAACT	GGCCCGCGCT	CGCGCCGGCC	ACACTCTCGT	ATGGGGGCGG	GGCGTCAATC	108840
	ACGCTATCAT	CTCCGTCATC	CCTGCATGCG	TGGGCATGCC	CAGCCCCCAA	CGCCATGGTG	108900
	GGGATTGCGG	GCTCAGAAGC	CTGCATGTCG	TGTGGTCCGT	CGTAGTCCAA	CGTGCTTCCC	108960
	CCACCCACCA	CACAGCCGGT	CCCCACGCCG	ACCACTAGAC	CGCAGACGTC	GCCCCAACCGA	109020
	GGTCCCCGTG	CACAGACCGC	GCCTTTTATA	GCCCCAGGGG	TTGCTAATTA	ACGCACGCAT	109080
35	GCAGACGCAA	TTTATTTTGC	TCCCCCGCGT	CCTCCCCCTC	CCCGCGTCCT	CCCCTCCCCG	109140
	TCCTCCCCCTC	CCCCGCGTCC	TCCCCCTCCC	CGCGTCCCTC	CCTCCCCCTGC	GCACACGTGA	109200
	TAGGTCTTTG	GAACCCGAGG	GGCGACGCGG	GGAAAGCGCG	CCCCCGCCCG	GCCGCGCGCG	109260
	GCCCCCGCCC	GGCCGCGCGG	CGCCCCCGCC	CGGCCGCGCG	CCCCCGCCCG	GCCGCGCGCG	109320
	GCCCCCGCCC	GGCCGCGCGG	CGCCCCCGCC	CGGCCGCGCG	GCGCCCCCGC	CCGGCCGCCC	109380
40	GCGTCGCGCC	GGCGCCCCCT	CCCGGCGCTT	CCGGGGCCTT	TCCTTCCCTT	CCCGCCGCGA	109440
	CCCCGGCCCC	GCCCCACCGC	CCCGCCCGGC	AGGGGGGCCC	CGGCGCCGCG	CAGAACACAC	109500
	AGACGAACAC	ACGGTGGCGA	TCTTTTCTTT	ACTTCGGCAG	ACCAGCGAGC	CCCGGCCCCG	109560
	GCCCGCGCCC	CGCCGCCACA	CCCACGGCAC	CCCCCCCCGCC	GCCCACCCCG	GGGTCCACAC	109620

	AGGAGCGCGC	GGGCGGCAGA	AACGCGGGCG	CGGCGGCGGT	CGGGGTGGGA	GTGGTGGTGG	109680
	GGGACACGAA	AACACACCCA	CGACACTCTC	CCCCCACCAC	GACCGCCGCC	GCGCCCCACC	109740
	GGCGGGATCG	CGGCGAGACG	CAGCCGGGCC	CCCCCCCACC	ACCCGCCAC	CCACCTACCC	109800
	CGCGCCCGCA	GCCTCCGGCA	GCACGCCGAC	CACCGCCGCC	ACCCCCAAA	CAGCCAAGGC	109860
5	GCGGTGGGGG	GCGTGGTGGT	GAACGATGGG	GGGAACACGG	GGGGGAGGGG	TCCGGGGCGA	109920
	GGCGGGCGGG	CGAAGGAAGG	GGGGGTGGTG	GCGGCGGGCG	TGGAAAGCGG	AAAAACGGAG	109980
	GATGGAAGGG	CAGAAGATGG	GGAGTCCCGA	TCCTCCTCCT	GCATCCCCTC	GCCTTCCATT	110040
	CTCCGGCCCT	CCGCGAGTCC	CGACGCCCCC	CCCCCGCCGC	CCGACGAAGG	AGACCCAAGC	110100
	ACCGCAGCCG	GAGAGGCCGA	GCGGGGAGTG	GCGGGCCGGG	CGGGAGGATG	GCGGAGAGAG	110160
10	AGAGAGAGAG	AGAGAGAGGG	GGGGGGGGGG	AGAGGGAAAG	CAACGGGAAA	GAGAGGCGCG	110220
	CGGAAAAGCA	GCAAGAGGGG	GGACGGGGCG	AGCCGGGCAG	AGTGCGGAGC	CCCCGGAGCC	110280
	CGCGGCCGCA	GCCGAGCAGC	GCCGCGGGCT	CCGGGGCCGG	GCCGGGCCGG	CAACCCCCCG	110340
	CGCCGGCCGC	GGCGGAGAGA	ACCCCTGTGT	CATTGTTTAC	GTGGCCGCGG	GCCAGCAGAC	110400
	GGGCCGCGGG	CCAGCAGACG	GGCCGCGGGC	CCAGCGGGCC	ACGCCCTCCG	CCGCATTAGG	110460
15	CCCCGCGGG	CATCCGGCGG	CCGGCCCCAC	GCCCTTCCAT	TAAACACTCC	CACGTTGGGG	110520
	GGGGGCGCGC	CAGCTGAGTG	CTCTGCGGTT	GCGGGCGCCG	TGCCCCGAGA	TCCATTAAGC	110580
	CGCCGGAGAG	CCCAGACCCC	GCCCGCGTGT	TGCTGTGGGC	ATTTCTGCTG	CGTCATCCCT	110640
	GTCTTTTATA	AACCGGGGGC	GCGGCAGCAA	CGAACACAGG	GGCCCCCGCG	CGATCGAGAG	110700
	GGACTCCGGA	GAAGGAAGGC	TGCTCCGCGC	ACCGGCGCGC	CCTTCTCCTC	TCCCTCCCT	110760
20	ACCTCCCCCT	CTCTTCCCCC	TTTFTTCCCC	CGCCTCCCGT	CTTCTTCCGC	GCCTCCGAGG	110820
	GTCCGCCTCT	GCCTCGGGGA	CCCCCGGGCG	GGCCGGGGCT	TGGCCGCCGA	GGTGCGCCCC	110880
	GGCCGGAGGG	GCCCCCGCAC	CTCGGCGGCC	GCCCCCTCCG	GCGCCGCGCG	TTCGCGAAAG	110940
	GCGCGAAAGG	GGCCCCCGGA	GGCTTTTTTC	GATTCCCGGC	CGGGGGTCCC	GGGTAGCCGC	111000
	CCGGCGCCCG	TGGGAAGGCG	TCCCCCGCCC	GGCGGTCTNG	NNNNGGCCCC	CGGCGGAGCG	111060
25	CGGGGGCCCC	GGGGCCCCGG	GCCGCGCCGG	CGGCGTTTCC	GCGTTCGGTT	TCTTCTCCCT	111120
	CCCGGCCGCC	CCGCTCCCGG	GCCCGACCTT	CGCCCCTTCC	CTTCTCCTCG	TCTTCCCCCG	111180
	TCCCGCCGCG	CCCCCTCCCT	CTTCTTCTCT	TCTCTCTGTC	TCGCTGTCTC	GCTCTCCTCA	111240
	CATTTCCCCC	CCCCCCCCCC	GGCCGCGGCC	GCCGCCCTCT	GCCCGCGTCC	CACCGAGACG	111300
	CCGCGCCGCG	TGAGCCGTCC	GCCGGGGGAC	CCAGGCTCCG	GGGGGGGGGC	GCGCCTGCGT	111360
30	GTGTCTCGTG	TGAGAGAGCG	CGCCCCCTCGA	ACGCCGCGCG	TTCTCGCAGG	TAGGTTTAGG	111420
	GTCGTACAGG	TGAGCTTCTG	CTGAGCGCGC	GGGAGAGGGG	GGGGCGGGCG	GAAGAGAGAA	111480
	GAGAGCAGGG	GTTGGGGGAA	AACTGTTCTT	CCTCCCCCTT	TCAAGAAACA	CGAGGCGGGG	111540
	GTCCAGAAA	GGGCAGGCAG	GTCAGCCGCA	CCGCCCGCGA	GCCAACCCGT	ATCCTTTTTT	111600
	TCTAGGTGTT	TTTGTTTTTG	TTTCTGTTTT	TGTTTGTTTT	GTTATTATTT	TCGCGGATCC	111660
35	GGCGTGTTCG	GATCCACCCC	CCCCTTCTCT	CTTCTCTTTC	CCTTCCACCC	ACCCCCGTTT	111720
	CCCCCCCCC	CCCGTGGTGT	CGTCCGGGGG	CGTCTGTCCC	AGGGGGGGCA	GGCGCGGGTC	111780
	GGGCCCATAC	GCCCACCGCC	CCCACGCGCC	GGTCACCCCC	CCCCCAACAA	CCCCAAAGGC	111840
	GCGTGCCCCG	CCACAGCCGT	GGGTGTGGCG	CCCGTCCCTT	TCCTCTACCG	CGTGGGCGCG	111900
	GGCGGGGGGG	TGGTGGTGGT	AGTGGTGGCG	GAAGGAAACG	GGCCGGGGGG	CCGGGGCCGC	111960
40	TAGGGAAAGG	TAGGCACGCG	CGCGGTGTGT	CGACTTG CAT	GCCCCGAAA	ACGCGTCTGT	112020
	TCGTGTGTGT	TCGTGGTGGG	CCGTGTTGTG	GTGGGCCGTG	TGGTGTGGTG	TGGTGTGGCG	112080
	AACGCGCGAG	CCCCCTCGCC	CCGATGGGAG	TCTCCCCGCA	GCCAGGGTAA	GGAGGGGCGG	112140
	GCGTGCGGG	CAGGTGTGCG	GGCGGGGTGG	GGTGTGTGCG	GTTGCATGCC	TCGGGTCTCC	112200



	TCTTCCTGCT	CCTCCTCCTT	TCTCCCAGCC	AGGGTGAGGA	GGGGCGGGCG	TGGCGGGCGG	112260
	GTGTGCGGGC	GGGGTGGGCG	CCGGGGCGGG	GGTGGGCACG	GGCGTAAGTG	CGGGTGCATG	112320
	CCTCGGGTCT	TCTCTTCTCC	CTCCTCCTTC	CTCCCACCCG	TCCCCGGGGG	CAGAGGGCGT	112380
	GCAATGCGTTG	TGATTCAACC	GCCCTCGCCC	CCGCCCCACT	TCCCCCCTC	TCTATCAAAG	112440
5	TTCCCTGGCC	CCTGGCTTCG	CGCCGGTGGT	GCGGCTGACC	CCCCCTCCT	CCCTCCCCGA	112500
	GCCAGGCGCC	CTCCCCTCC	TGCCCCACCAC	CCCCCGGGTC	TGGCCGGCCA	GACGTGCGTG	112560
	CTCTGCACGA	TCGGGCCCCC	CTCCCTGTCA	ACACGGACAC	ACTCTTTTTT	TACCCGCCAG	112620
	CCCGCCCACC	CACCAAGACA	GGGAGCCAGA	ACGACGGCCG	GGCCCCGGCT	CTGTTCTATG	112680
	ATAAAGACCA	ACAGGCCTCG	GGGGTGGGGG	CGGCTTCTCG	TGCCCGCCAA	GGAGACGCCC	112740
10	GCCCCGACCA	CCCTCGCAGC	GCAGGCGCAG	GCCGCGACCC	CTCTCTGCTC	TTTGGAGGGA	112800
	GCCGGGGGCG	CGACGACGCG	GCGCCCCCGG	CTCCTTCACA	CGGTCTTCT	GCGCGGTGCG	112860
	CCTCCCGCCG	GGGCTCGGGC	CCCGGGCGCT	GGGCGCGGG	CCGAGTGCG	CGATGGACGG	112920
	GTAGCGCCCC	CAGAGCTCGC	AGCACCGGGA	CCGCGGAATG	CACCTGTTCT	GCCAGTGCCC	112980
	CCTGACGGAC	GGGCAGGACC	TGTACCTCTG	CCCGGTGTAT	CCCCGGATGC	ACCAGGAGCA	113040
15	CCTGGTCTGC	CCCTTGCACC	GCCTGGACGA	CGCCCGGCGC	CGGGGGCGCA	CCTCGGCGGC	113100
	GTGGGACGAG	GGGCTCGTGC	GCGCGTTGAC	GACTCCGGG	GGGCTGATGG	GCTGCGGGGG	113160
	GCGCAGCCTC	ACCTTGTCGG	AGACCTACTG	GGGCCACCCG	TTGTACGAGA	AACTGGTCCC	113220
	GTGGGACCAC	CCGCGCGACC	TGAAGGTGCC	GGAGGCCAGC	GCGGTGGGCA	CCAGAGCCCT	113280
	CGTCCCGCGC	GGGCGCGGCC	GGCCGCTGCG	GGGGCGCCCG	GTGCCCCCTCA	TCCCCCTCGA	113340
20	TTGTGAGCCG	AACGACGGGC	TTCCCTTCGG	CGGGGGGTGG	CCTGGTGGCC	GGCTCCGCGG	113400
	AGCCCCCGTC	CCCCCTCCACC	CCCCCCCCCC	TTCTGCCCCCT	CCTCTGTCCCT	TCACCCCCAC	113460
	CCTCACCCCC	CCCTGCCTGT	GCCGGGGCTT	GTCGTGTGT	GTGGTCGTAA	AACAATACCT	113520
	GAAAGACCGG	AACAACCTTT	GAATCCTTT	TTTTTGAAAT	ATAAATATTT	TTAAATGTT	113580
	ATTTCAAAC	ACTACGAAAA	CTGTGTGAAA	CAACAACCGG	AACTACGTC	GAGGGGGCGC	113640
25	GTCCCCCGGG	CCCCTACCCC	CCCCCTTTC	CCTCCTCCTC	CTCCCCGCCC	TCCGCTCCT	113700
	CCTCCTCCGC	CTCCTCCTCC	TCCGCTCCT	CCTCCTCCGC	CTCCTCCTCC	TCCGCTCCT	113760
	CCTCCTCCGC	CTCCTCCTCC	TCCGCTCCT	CCTCCTCCGC	CGCCGCTGGC	GCCGGACCT	113820
	GCTGCTCTTG	CGGCTGCCCC	CGCGCCGCGG	GCGCCTGCGG	CCCCGCTCGC	CGGGCACCGG	113880
	CGCCAGCGGG	CTCAGGCTCA	GGCCCCGGGC	CGCGCCGCGG	CGGGAGAACC	GGGGGTGGGG	113940
30	GACCCCCCGC	TCCCCGCTCG	CGCCCCGCGG	CCTCCTTCTC	CGCCTCCTGC	TCCGGCGCCC	114000
	CGGGCTCAGG	CTGGGCGCGG	AGAAGGCCCC	CGCCCGGAGA	GGAGGAGGCG	CGTCCACAGG	114060
	AGCCCGGGGC	CCCCCTCCTC	AGACGGTGTG	AGCAGCCCCG	CGCGCCGCGC	GGGGGCGCGC	114120
	CGGCAGCGGG	GCGCGCAGGC	CTCAGGCGGG	GCGCGGCGGC	GGCGGGGGCA	CCACAGACGC	114180
	TCGCGCCTGC	GCCGGCCCCG	GCGCGGCGGG	CGGCACGGCC	ACCTGCGCGT	GGCGCGCGGG	114240
35	GCCAGCGCGT	ACTGGGTCCG	AGTCTGGCTG	TGGGTTCGTG	TCTCAGACCC	GGCCCGTCCG	114300
	CGCTGGCTGC	GCGCGCCCAG	CCCTCCCGGC	CGCGCCTCCT	CTCCTGGGCC	CCAGGGGGCG	114360
	CCGTGGTTGT	GGGGGCCACG	GCGGGGGGTG	CGGCGCCTCC	CCCGCCGTCT	GTTTCTCTCT	114420
	GCCGGGCCCC	GGGCGCCCCG	CCGCGCCTCT	GCCGCCCCCT	CTCAGCGACT	ACTGATACCC	114480
	CCCGAGGACC	CGGCGCGCCC	CGACAGAGCG	CCCCCGCAG	GACGGGAGGC	GGCGGCGCCG	114540
40	CAGAAGCGGG	TGGGCGGCGC	GGACGCGCGC	GGGGGGCGGC	CGGCGTCCCC	CTTCTCTCCG	114600
	GTGAGAGCCG	TGCTGCCGCG	GCTGCCGTCC	CGGCGGGGGT	ATGCGGTTCC	CGCGTTAATT	114660
	GGGAGTGATT	TCCCTTGTTT	TCGACCTCGA	GGTGGCGCCA	CCGCCGGCGA	GATCTTGATC	114720
	ACCTAGGGGG	CCCACGTCC	TTAAGTACT	GGGTCTAGGG	TGGGGGCGGG	CGTTGCCCCG	114780

	CGGCGGCGAC	GACGACGAGG	CGCCCCGCGG	TCCCCCGCGG	CCAGCCCAGC	GCCGCCCCGAC	114840
	CCTCCAAGGC	GCCCAGCGGG	GGCGTGGCGG	CGGGGGCGCG	GCCCCGCGAG	AAGCCCCCGG	114900
	CCCCGCCCTGC	ATCAGGCGAC	GTCTCCCTCT	GTCTCTGCCC	TTGGGGGCCA	ATCACGGGCT	114960
	GGGGGCGGGC	TGGGGGCGGG	TCACGGGCTG	GGGGCGGGTC	ACGGGCTGGG	GGCGGGTCAC	115020
5	GGGCTGGGGG	CGGGTCACGG	GCTGGGGGCG	GGCGGGAGTG	GCAGCCGGTC	CAGTAGCAGG	115080
	AGCAGCAGGC	ACGGCCCCGT	CCCCCCCCAC	CCGTGTCCCC	GCGCCTGGCA	CACAGGGGGG	115140
	TCGCTGTCCC	TCGCGCCCCG	GCAGGCGCCC	AACGGGCAGG	TCTATTTTCT	GTGCCGGCAC	115200
	GGCCTGGCGT	GCCGGCGGAG	CCGGAGGTGC	GCCCAGGCCC	CCAGCAAGTG	ATAGCCCTAC	115260
	CACGACTTGC	TGGGCGACCG	CCAGTGCGGG	TGATAGTCCA	TGCGGTGGCC	CCACAACGTG	115320
10	TCCCCGTGTC	ACAACGCGTT	GCCTTAGGTC	CAGAAGTACG	TGCCCTACGT	CTTCCCCACG	115380
	TCCGTCCCTT	TTGAGACCGT	CGCGTCCCCG	CCCCGCTAGA	GCAGGCACGT	GTGCCGTGTG	115440
	TGCAGCGGGG	GGGGAGGGCG	AAGGCGAAGG	AGGAGTGGGT	GCCCGGGTGG	GGGCGTCTTA	115500
	GGGACGCGCA	GCCGCCCCGA	CCCCGACGGG	ACCGCGAGCC	GGCCCCCGGC	CCGGCCCCCG	115560
	CACCGGCGCA	GGTAGTCCGG	GCGGAGCTTG	TAGAGGCACA	GGCACGACGG	GCGGAGCCTC	115620
15	CACCTCAGCG	CCACTTCCAG	CAGCAGTCTC	TAAGGTGGA	GCCAGAGGAG	GAGGCTCAGC	115680
	GACGACCGCT	CGGTGACGTA	CAGCAACTCG	TAGGGGTCC	GCACGCCCCG	CCGCCCCGACG	115740
	AACGTGTTCT	TTGCCCCCCC	CTAAATCTCC	CGCGCCCCGC	ACTCCGCCCT	GGGGGCACGG	115800
	CACAGGGGGC	ACAGGGAGGG	AGTGGGGCGG	GGGGGCGGGC	GACGAAAAAC	AAGCCTTCCC	115860
	CCCCCTCTTC	CCCAGGCATT	GGTTTCCACC	AGACGCAGGA	AACCTAAGGC	TGGGGAGCAG	115920
20	AGGGGGGGGG	ACAGGGGGCG	AGAGCCCCGAG	TCCCGAGGGA	CGGAGGGAGC	GGGGGGGTTT	115980
	CCCAGCCCCC	CGCCGCGTGC	CGGGTGCCCC	CAGGGGGCTG	GCGAATTTCG	CCGGCCCCCA	116040
	GCCGGGGCAG	TTCGAGGGG	CGGGGGCTCG	GGTGGCGGGC	GCTGGTGGGG	GTTGGGCGTC	116100
	GGCCCAACAG	GCCCCCTTTC	CCCCCCGGAC	TCTGGGCCCC	CAGCGGGAGA	GTGGCACGGC	116160
	CCCCAGACGG	CGCCGCCGGC	GAGCCCCGGC	CCCAGGCGGG	CCCTCGAGCA	CGGCCCGGCC	116220
25	CCAAGGTACT	CGGCCCCATC	CCATCTGAGC	TCTGCCGCCG	GGCGCCAGAG	AGAGAACGGC	116280
	CCACAATCAG	AGACAGAGAG	GCCCAGAGGA	GGAGGCGGGC	CCGCGGGCGA	GGCAGCGAGC	116340
	GTCACGGCCC	CACGCTTACG	CCGGGCTGGC	AGTGTGCCCC	GACGGAATAT	GGGCCGCGGA	116400
	TAGGTGAGGG	GGTTTCCCCG	CCGTAAATGC	TAAGGGGGTT	ATCGGCGCGC	GGGGCCGCCC	116460
	CCGCTTCCCT	CCCTTAGGGG	GGGAGAGCCC	CGCCGGGGCA	GGGGCCCTTG	GTTGGCCAC	116520
30	ATGAGGTTCT	TGGGGTAATC	GTACGCGCGG	GGGGGCGGCT	GCGTCTACCC	TCAGGGGGGC	116580
	CGCGGGGCGG	CCGCGCCGGG	ACTCACCACG	GGCGGGGGCC	CCTCTTTAAG	TAATCGTATG	116640
	ATCCTTCGGG	TCCCCTGGTT	ATCCCCGGCT	AGTCGGGTGG	GTGGGCGGCC	GCGCGCTCCG	116700
	AGACGCACAA	GACGGTTCTT	TCATTAGTCG	TATTGGGCCT	TGGGGCTCCC	TCATTAATGC	116760
	GCCCCTCGCT	CCCCGGCAGG	CTTGCAAAAA	TTAATGGTAT	TCGCCCTTAC	CGCCGGGCAA	116820
35	TTTTCGACGA	TTAATGGCGC	TCGCCCTTGC	GGCCGGGTAA	TTTTCACGA	TTAATGGTAC	116880
	TCGCCCTAC	CGCCGGCCCT	GGCGGATAAT	TTTCAAAGAT	TAATGGTATG	GCCCTTCGGC	116940
	CGCGCCCCGC	CAGCGGCCCC	GCCTCAGGCC	CGGGCGCGCC	GCCGCGCGCC	AACCGGCGGC	117000
	GGCGGGGGAC	CCCCCGCGCC	TCGCCGCCCC	GCCGCGGCC	GGGAGCGCCT	ATATATGCGC	117060
	CCCGAGGGTA	GCAGAGAAGC	CTCTCGCCGG	AGCGGTCTG	GAAGCCTCGA	GGCCCCGAGG	117120
40	CGGCCGGCTC	CGGCGGGAGC	GGCCAAGTTG	GGATCTGGCG	GGCTGCCGGG	CCCGGGCGCC	117180
	GCCGCTCTCT	GGGCGCGCGG	CGGCGGCGGC	GGA			117213

(2) INFORMATION FOR SEQ ID NO:218:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 180 amino acids

(B) TYPE: amino acid

5 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:

Met Arg Thr Pro Ala Asp Asp Val Ser Trp Arg Tyr Glu Ala Pro Ser  
 1 5 10 15  
 Val Ile Asp Tyr Ala Arg Ile Asp Gly Ile Phe Leu Arg Tyr His Cys  
 15 20 25 30  
 Pro Gly Leu Asp Thr Phe Leu Trp Asp Arg His Ala Gln Arg Ala Tyr  
 35 40 45  
 Leu Val Asn Pro Phe Leu Phe Ala Ala Gly Phe Leu Glu Asp Leu Ser  
 50 55 60  
 20 His Ser Val Phe Pro Ala Asp Thr Gln Glu Thr Thr Thr Arg Arg Ala  
 65 70 75 80  
 Leu Tyr Lys Glu Ile Arg Asp Ala Leu Gly Ser Arg Lys Gln Ala Val  
 85 90 95  
 Ser His Ala Pro Val Arg Ala Gly Cys Val Asn Phe Asp Tyr Ser Arg  
 25 100 105 110  
 Thr Arg Arg Cys Val Gly Arg Arg Asp Leu Arg Pro Ala Asn Thr Thr  
 115 120 125  
 Ser Thr Trp Glu Pro Pro Val Ser Ser Asp Asp Glu Ala Ser Ser Gln  
 130 135 140  
 30 Ser Lys Pro Leu Ala Thr Gln Pro Pro Val Leu Ala Leu Ser Asn Ala  
 145 150 155 160  
 Pro Pro Arg Arg Val Ser Pro Thr Arg Gly Arg Arg Arg His Thr Arg  
 165 170 175  
 Leu Arg Arg Asn  
 35 180

## (2) INFORMATION FOR SEQ ID NO:219:

## (i) SEQUENCE CHARACTERISTICS:

40 (A) LENGTH: 334 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:

5  
Met Lys Arg Ala Arg Ser Arg Ser Pro Ser Pro Pro Ser Arg Pro Ser  
1 5 10 15  
Ser Pro Phe Arg Thr Pro Pro His Gly Gly Ser Pro Arg Arg Glu Val  
20 25 30  
10 Gly Ala Gly Ile Leu Ala Ser Asp Ala Thr Ser His Val Cys Ile Ala  
35 40 45  
Ser His Pro Gly Ser Gly Ala Gly Tyr Pro Thr Arg Leu Ala Ala Gly  
50 55 60  
Ser Ala Val Gln Arg Arg Arg Pro Arg Gly Cys Pro Pro Gly Val Met  
15 65 70 75 80  
Phe Ser Ala Ser Thr Thr Pro Glu Gln Pro Leu Gly Leu Ser Gly Asp  
85 90 95  
Ala Thr Pro Pro Leu Pro Thr Ser Val Pro Leu Asp Trp Ala Ala Phe  
100 105 110  
20 Arg Arg Ala Phe Leu Ile Asp Asp Ala Trp Arg Pro Leu Leu Glu Pro  
115 120 125  
Glu Leu Ala Asn Pro Leu Thr Ala Arg Leu Leu Ala Glu Tyr Asp Arg  
130 135 140  
Arg Cys Gln Thr Glu Glu Val Leu Pro Pro Arg Glu Asp Val Phe Ser  
25 145 150 155 160  
Trp Thr Arg Tyr Cys Thr Pro Asp Asp Val Arg Val Val Ile Ile Gly  
165 170 175  
Gln Asp Pro Tyr His His Pro Gly Gln Ala His Gly Leu Ala Phe Ser  
180 185 190  
30 Val Arg Ala Asp Val Pro Val Pro Pro Ser Leu Arg Asn Val Leu Ala  
195 200 205  
Ala Val Lys Asn Cys Tyr Pro Asp Ala Arg Met Ser Gly Arg Gly Cys  
210 215 220  
Leu Glu Lys Trp Ala Arg Asp Gly Val Leu Leu Leu Asn Thr Thr Leu  
35 225 230 235 240  
Thr Val Lys Arg Gly Ala Ala Ala Ser His Ser Lys Leu Gly Trp Asp  
245 250 255  
Arg Phe Val Gly Gly Val Val Arg Arg Leu Ala Ala Arg Arg Pro Gly  
260 265 270  
40 Leu Val Phe Met Leu Trp Gly Ala His Ala Gln Asn Ala Ile Arg Pro  
275 280 285  
Asp Pro Arg Gln His Tyr Val Leu Lys Phe Ser His Pro Ser Pro Leu  
290 295 300

Ser Lys Val Pro Phe Gly Thr Cys Gln His Phe Leu Ala Ala Asn Arg  
 305 310 315 320  
 Tyr Leu Glu Thr Arg Asp Ile Met Pro Ile Asp Trp Ser Val  
 325 330

5

(2) INFORMATION FOR SEQ ID NO:220:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 231 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: peptide

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:

Met Val Lys Ser Arg Val Ser Tyr Arg Ser Val Met Ser Gly Val Gly  
 1 5 10 15  
 20 Glu Glu Arg Val Pro Ser Ala Phe Thr Ile Leu Ala Ser Trp Gly Trp  
 20 25 30  
 Thr Phe Ala Pro Gln Asn His Asp Pro Gly Asp Asn Thr Thr Pro Ile  
 35 40 45  
 Glu Ser Ile Ala Gly Thr Ala Pro Asp Ala His Val Gly Pro Leu Asp  
 25 50 55 60  
 Gly Glu Pro Asp Arg Asp Ala Ile Ser Pro Leu Thr Ser Ser Val Ala  
 65 70 75 80  
 Gly Asp Pro Pro Gly Ala Asp Gly Pro Tyr Val Thr Phe Asp Thr Leu  
 85 90 95  
 30 Phe Met Val Ser Ser Ile Asp Glu Leu Gly Arg Arg Gln Leu Thr Asp  
 100 105 110  
 Thr Ile Arg Lys Asp Leu Arg Leu Ser Leu Ala Lys Phe Ser Ile Ala  
 115 120 125  
 Cys Thr Lys Thr Ser Ser Phe Ser Gly Thr Ala Ala Arg Gln Arg Lys  
 35 130 135 140  
 Arg Gly Ala Pro Pro Gln Arg Thr Cys Val Pro Arg Ser Asn Lys Ser  
 145 150 155 160  
 Leu Gln Met Phe Val Leu Cys Lys Arg Ala Asn Ala Ala Gln Val Arg  
 165 170 175  
 40 Glu Gln Leu Arg Ala Val Ile Arg Ser Arg Lys Pro Arg Lys Tyr Tyr  
 180 185 190  
 Thr Arg Ser Ser Asp Gly Arg Leu Cys Pro Ala Val Pro Val Phe Val  
 195 200 205

His Glu Phe Val Ser Ser Glu Pro Met Arg Leu His Arg Asp Asn Val  
 210 215 220  
 Met Leu Ser Thr Glu Pro Asp..  
 225 230

5

(2) INFORMATION FOR SEQ ID NO:221:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 199 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

Met Gly Asn Pro Gln Thr Thr Ile Ala Tyr Ser Leu His His Pro Arg  
 1 5 10 15  
 20 Ala Ser Leu Thr Ser Ala Leu Pro Asp Ala Ala Gln Val Val His Val  
 20 25 30  
 Phe Glu Ser Gly Thr Arg Ala Val Leu Thr Arg Gly Arg Ala Arg Gln  
 35 40 45  
 Asp Arg Leu Pro Arg Gly Gly Val Val Ile Gln His Thr Pro Ile Gly  
 25 50 55 60  
 Leu Leu Val Ile Ile Asp Cys Arg Ala Glu Phe Cys Ala Tyr Arg Phe  
 65 70 75 80  
 Ile Gly Arg Ala Ser Thr Gln Arg Leu Glu Arg Trp Trp Asp Ala His  
 85 90 95  
 30 Met Tyr Ala Tyr Pro Phe Asp Ser Trp Val Ser Ser Ser His Gly Glu  
 100 105 110  
 Ser Val Arg Ser Ala Thr Ala Gly Ile Leu Thr Val Val Trp Thr Pro  
 115 120 125  
 Asp Thr Ile Tyr Ile Thr Ala Thr Ile Tyr Gly Thr Ala Pro Glu Ala  
 35 130 135 140  
 Arg Cys Asp Asn Ala Pro Leu Asp Val Arg Pro Thr Thr Pro Pro Ala  
 145 150 155 160  
 Pro Val Ser Pro Thr Ala Gly Glu Phe Pro Ala Asn Thr Thr Asp Leu  
 165 170 175  
 40 Leu Val Glu Val Leu Arg Glu Ile Gln Ile Ser Pro Thr Leu Asp Asp  
 180 185 190  
 Ala Asp Pro Thr Pro Gly Thr  
 195

## (2) INFORMATION FOR SEQ ID NO:222:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 877 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 10 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

```

Met Ala Ala Ser Gly Gly Glu Gly Ser Arg Asp Val Arg Ala Pro Gly
1      5      10      15
Pro Pro Pro Gln Gln Pro Gly Ala Arg Pro Ala Val Arg Phe Arg Asp
      20      25      30
Glu Ala Phe Leu Asn Phe Thr Ser Met His Gly Val Gln Pro Ile Ile
      35      40      45
20 Ala Arg Ile Arg Glu Leu Ser Gln Gln Gln Leu Asp Val Thr Gln Val
      50      55      60
Pro Arg Leu Gln Trp Phe Arg Asp Val Ala Ala Leu Glu Val Pro Thr
65      70      75      80
Gly Leu Pro Leu Arg Glu Phe Pro Phe Ala Ala Tyr Leu Ile Thr Gly
25      85      90      95
Asn Ala Gly Ser Gly Lys Ser Thr Cys Val Gln Thr Leu Asn Glu Val
      100     105     110
Leu Asp Cys Val Val Thr Gly Ala Thr Arg Ile Ala Ala Gln Asn Met
      115     120     125
30 Tyr Val Lys Leu Ser Gly Ala Phe Leu Ser Arg Pro Ile Asn Thr Ile
      130     135     140
Phe His Glu Phe Gly Phe Arg Gly Asn His Val Gln Ala Gln Leu Gly
145     150     155     160
Gln His Pro Tyr Thr Leu Ala Ser Ser Pro Ala Ser Leu Glu Asp Leu
35      165     170     175
Gln Arg Arg Asp Leu Thr Tyr Tyr Trp Glu Val Ile Leu Asp Ile Thr
      180     185     190
Lys Arg Ala Ala His Gly Gly Glu Asp Ala Arg Asn Glu Phe His Ala
      195     200     205
40 Leu Thr Ala Leu Glu Gln Thr Leu Gly Leu Gly Gln Gly Ala Leu Thr
      210     215     220
Arg Leu Ala Ser Val Thr His Gly Ala Leu Pro Ala Phe Thr Arg Ser
225     230     235     240

```

	Asn	Ile	Ile	Val	Ile	Asp	Glu	Ala	Gly	Leu	Leu	Gly	Arg	His	Leu	Leu	
					245					250					255		
	Thr	Thr	Val	Val	Tyr	Cys	Trp	Trp	Met	Ile	Asn	Ala	Leu	Tyr	His	Thr	
				260					265					270			
5	Pro	Gln	Tyr	Ala	Gly	Arg	Leu	Arg	Pro	Val	Leu	Val	Cys	Val	Gly	Ser	
			275					280					285				
	Pro	Thr	Gln	Thr	Ala	Ser	Leu	Glu	Ser	Thr	Phe	Glu	His	Gln	Lys	Leu	
			290				295					300					
	Arg	Cys	Ser	Val	Arg	Gln	Ser	Glu	Asn	Val	Leu	Thr	Tyr	Leu	Ile	Cys	
10	305					310					315				320		
	Asn	Arg	Thr	Leu	Arg	Glu	Tyr	Thr	Arg	Leu	Ser	His	Ser	Trp	Ala	Ile	
				325					330					335			
	Phe	Ile	Asn	Asn	Lys	Arg	Cys	Val	Glu	His	Glu	Phe	Gly	Asn	Leu	Met	
				340					345					350			
15	Lys	Val	Leu	Glu	Tyr	Gly	Leu	Pro	Ile	Thr	Glu	Glu	His	Met	Gln	Phe	
			355					360					365				
	Val	Asp	Arg	Phe	Val	Val	Pro	Glu	Ser	Tyr	Ile	Thr	Asn	Pro	Ala	Asn	
			370				375					380					
	Leu	Pro	Gly	Trp	Thr	Arg	Leu	Phe	Ser	Ser	His	Lys	Glu	Val	Ser	Ala	
20	385					390					395				400		
	Tyr	Met	Ala	Lys	Leu	His	Ala	Tyr	Leu	Lys	Val	Thr	Arg	Glu	Gly	Glu	
				405					410					415			
	Phe	Val	Val	Phe	Thr	Leu	Pro	Val	Leu	Thr	Phe	Val	Ser	Val	Lys	Glu	
				420					425					430			
25	Phe	Asp	Glu	Tyr	Arg	Arg	Leu	Thr	Gln	Gln	Pro	Thr	Leu	Thr	Met	Glu	
			435				440					445					
	Lys	Trp	Ile	Thr	Ala	Asn	Ala	Ser	Arg	Ile	Thr	Asn	Tyr	Ser	Gln	Ser	
			450				455					460					
	Gln	Asp	Gln	Asp	Ala	Gly	His	Val	Arg	Cys	Glu	Val	His	Ser	Lys	Gln	
30	465					470					475				480		
	Gln	Leu	Val	Val	Ala	Arg	Asn	Asp	Ile	Thr	Tyr	Val	Leu	Asn	Ser	Gln	
				485					490					495			
	Val	Ala	Val	Thr	Ala	Arg	Leu	Arg	Lys	Met	Val	Phe	Gly	Phe	Asp	Gly	
				500					505					510			
35	Thr	Phe	Arg	Thr	Phe	Glu	Ala	Val	Leu	Arg	Asp	Asp	Ser	Phe	Val	Lys	
			515					520					525				
	Thr	Gln	Gly	Glu	Thr	Ser	Val	Glu	Phe	Ala	Tyr	Arg	Phe	Leu	Ser	Arg	
			530				535					540					
	Leu	Met	Phe	Gly	Gly	Leu	Ile	His	Phe	Tyr	Asn	Phe	Leu	Gln	Arg	Pro	
40	545					550					555				560		
	Gly	Leu	Asp	Ala	Thr	Gln	Arg	Thr	Leu	Ala	Tyr	Gly	Arg	Leu	Gly	Glu	
				565					570					575			
	Leu	Thr	Ala	Glu	Leu	Leu	Ser	Leu	Arg	Arg	Asp	Ala	Ala	Gly	Ala	Ser	



	580	585	590
	Ala Thr Arg Ala Ala Asp Thr Ser Asp Arg Ser Pro Gly Glu Arg Ala		
	595	600	605
	Phe Asn Phe Lys His Leu Gly Pro Arg Asp Gly Gly Pro Asp Asp Phe		
5	610	615	620
	Pro Asp Asp Asp Leu Asp Val Ile Phe Ala Gly Leu Asp Glu Gln Gln		
	625	630	635
	Leu Asp Val Phe Tyr Cys His Tyr Ala Leu Glu Glu Pro Glu Thr Thr		
	645	650	655
10	Ala Ala Val His Ala Gln Phe Gly Leu Leu Lys Arg Ala Phe Leu Gly		
	660	665	670
	Arg Tyr Leu Ile Leu Arg Glu Leu Phe Gly Glu Val Phe Glu Ser Ala		
	675	680	685
	Pro Phe Ser Thr Tyr Val Asp Asn Val Ile Phe Arg Gly Cys Glu Leu		
15	690	695	700
	Leu Thr Gly Ser Pro Arg Gly Gly Leu Met Ser Val Gln Thr Asp Asn		
	705	710	715
	Tyr Thr Leu Met Gly Tyr Thr Tyr Thr Arg Val Phe Ala Phe Ala Glu		
	725	730	735
20	Glu Leu Arg Arg Arg His Ala Thr Ala Gly Val Ala Glu Phe Leu Glu		
	740	745	750
	Glu Ser Pro Leu Pro Tyr Ile Val Leu Arg Asp Gln His Gly Phe Met		
	755	760	765
	Ser Val Val Asn Thr Asn Ile Ser Glu Phe Val Glu Ser Ile Asp Ser		
25	770	775	780
	Thr Glu Leu Ala Met Ala Ile Asn Ala Asp Tyr Gly Ile Ser Ser Lys		
	785	790	795
	Leu Ala Met Thr Ile Thr Arg Ser Gln Gly Leu Ser Leu Asp Lys Val		
	805	810	815
30	Ala Ile Cys Phe Thr Pro Gly Asn Leu Arg Leu Asn Ser Ala Tyr Val		
	820	825	830
	Ala Met Ser Arg Thr Thr Ser Ser Glu Phe Leu His Met Asn Leu Asn		
	835	840	845
	Pro Leu Arg Glu Arg His Glu Arg Asp Asp Val Ile Ser Glu His Ile		
35	850	855	860
	Leu Ser Ala Leu Arg Asp Pro Asn Val Val Ile Val Tyr		
	865	870	875

(2) INFORMATION FOR SEQ ID NO:223:

40

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 292 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

	Met	Ala	Asp	Pro	Thr	Pro	Ala	Asp	Glu	Gly	Thr	Ala	Ala	Ala	Ile	Leu
	1				5					10					15	
10	Lys	Gln	Ala	Ile	Ala	Gly	Asp	Arg	Ser	Leu	Val	Glu	Val	Ala	Glu	Gly
				20					25					30		
	Ile	Ser	Asn	Gln	Ala	Leu	Leu	Arg	Met	Ala	Cys	Glu	Val	Arg	Gln	Val
			35					40					45			
	Ser	Asp	Arg	Gln	Pro	Arg	Phe	Thr	Ala	Thr	Ser	Val	Leu	Arg	Val	Asp
15		50					55				60					
	Val	Thr	Pro	Arg	Gly	Arg	Leu	Arg	Phe	Val	Leu	Asp	Gly	Ser	Ser	Asp
	65					70				75					80	
	Asp	Ala	Tyr	Val	Ala	Ser	Glu	Asp	Tyr	Phe	Lys	Arg	Cys	Gly	Asp	Gln
				85						90					95	
20	Pro	Tyr	Gly	Phe	Ala	Val	Val	Val	Leu	Thr	Ala	Asn	Glu	Asp	His	Val
				100					105					110		
	His	Ser	Leu	Ala	Val	Pro	Pro	Leu	Val	Leu	Leu	His	Arg	Leu	Ser	Leu
				115				120					125			
	Phe	Arg	Pro	Thr	Asp	Leu	Arg	Asp	Phe	Glu	Leu	Val	Cys	Leu	Leu	Met
25		130					135					140				
	Tyr	Leu	Glu	Asn	Cys	Pro	Arg	Ser	His	Ala	Thr	Pro	Ser	Leu	Phe	Val
	145				150					155					160	
	Lys	Val	Ser	Ala	Trp	Leu	Gly	Val	Val	Ala	Arg	His	Asp	Phe	Glu	Arg
				165						170					175	
30	Val	Arg	Cys	Leu	Leu	Leu	Arg	Ser	Cys	His	Trp	Ile	Leu	Asn	Thr	Leu
				180					185					190		
	Met	Cys	Met	Ala	Gly	Val	Lys	Pro	Phe	Asp	Asp	Glu	Leu	Val	Leu	Pro
			195					200					205			
	His	Trp	Tyr	Met	Ala	His	Tyr	Leu	Leu	Ala	Asn	Asn	Pro	Pro	Pro	Val
35		210					215					220				
	Leu	Ser	Ala	Leu	Phe	Cys	Ala	Thr	Pro	Gln	Ser	Ser	Ala	Leu	Gln	Leu
	225					230					235				240	
	Pro	Gly	Pro	Val	Pro	Arg	Thr	Asp	Cys	Val	Ala	Tyr	Asn	Pro	Ala	Gly
				245						250					255	
40	Val	Met	Gly	Ser	Cys	Trp	Lys	Ser	Lys	Asp	Leu	Arg	Ser	Ala	Leu	Val
				260					265					270		
	Tyr	Trp	Trp	Leu	Ser	Gly	Ser	Pro	Lys	Arg	Arg	Thr	Ser	Ser	Leu	Phe
				275					280					285		

Tyr Arg Phe Cys  
290

(2) INFORMATION FOR SEQ ID NO:224:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 734 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

10

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

15

	Met	Glu	Ala	Pro	Gly	Ile	Val	Trp	Val	Glu	Glu	Ser	Val	Ser	Ala	Ile	
	1				5				10					15			
	Thr	Leu	Tyr	Ala	Val	Trp	Leu	Pro	Pro	Arg	Thr	Arg	Asp	Cys	Leu	His	
				20				25						30			
20	Ala	Leu	Leu	Tyr	Leu	Val	Cys	Arg	Asp	Ala	Ala	Gly	Glu	Ala	Arg	Ala	
			35				40					45					
	Arg	Phe	Ala	Glu	Val	Ser	Val	Gly	Ser	Ser	Asp	Leu	Gln	Asp	Phe	Tyr	
		50				55					60						
	Gly	Ser	Pro	Asp	Val	Ser	Ala	Ala	Gly	Ala	Val	Ala	Ala	Ala	Arg	Ala	
25	65				70					75				80			
	Ala	Pro	Ala	Asp	Leu	Glu	Pro	Leu	Gly	Asp	Pro	Thr	Leu	Trp	Arg	Ala	
				85				90					95				
	Leu	Tyr	Ala	Cys	Val	Leu	Ala	Ala	Leu	Glu	Arg	Gln	Thr	Gly	Pro	Val	
			100				105					110					
30	Phe	Val	Pro	Leu	Arg	Leu	Gly	Trp	Asp	Pro	Gln	Thr	Gly	Leu	Val	Val	
			115				120					125					
	Arg	Val	Glu	Arg	Ala	Ser	Trp	Gly	Pro	Pro	Ala	Ala	Pro	Arg	Ala	Ala	
			130				135					140					
	Leu	Leu	Asp	Val	Glu	Ala	Lys	Val	Asp	Val	Asp	Pro	Leu	Ala	Ala	Arg	
35	145				150					155				160			
	Val	Ala	Glu	His	Pro	Gly	Ala	Arg	Leu	Ala	Trp	Ala	Arg	Leu	Ala	Ala	
				165					170				175				
	Ile	Arg	Asp	Ser	Pro	Gln	Cys	Ala	Ser	Ser	Ala	Ser	Leu	Ala	Val	Thr	
			180						185				190				
40	Ile	Thr	Thr	Arg	Thr	Ala	Arg	Phe	Ala	Arg	Glu	Tyr	Thr	Thr	Leu	Ala	
			195				200					205					
	Phe	Pro	Pro	Thr	Ser	Lys	Glu	Gly	Ala	Phe	Ala	Asp	Leu	Val	Glu	Val	
			210				215					220					

	Cys	Glu	Val	Gly	Leu	Arg	Pro	Arg	Gly	His	Pro	Gln	Arg	Val	Thr	Ala
	225					230					235					240
	Arg	Val	Leu	Leu	Pro	Arg	Gly	Tyr	Asp	Tyr	Phe	Val	Ser	Ala	Gly	Asp
					245					250						255
5	Gly	Phe	Ser	Ala	Pro	Ala	Leu	Val	Phe	Arg	Gln	Trp	His	Thr	Thr	Val
				260					265					270		
	His	Ala	Ala	Pro	Gly	Ala	Pro	Val	Phe	Ala	Phe	Leu	Gly	Pro	Gly	Phe
		275						280					285			
	Glu	Val	Arg	Gly	Gly	Pro	Val	Gln	Tyr	Phe	Ala	Val	Leu	Gly	Phe	Pro
10		290					295					300				
	Gly	Trp	Pro	Thr	Phe	Thr	Val	Pro	Ala	Ala	Ala	Ala	Ala	Glu	Ser	Ala
	305				310					315						320
	Arg	Asp	Leu	Val	Arg	Gly	Ala	Ala	Ala	Thr	His	Ala	Ala	Cys	Leu	Gly
					325					330						335
15	Ala	Trp	Pro	Ala	Val	Gly	Ala	Arg	Val	Val	Leu	Pro	Pro	Arg	Ala	Trp
					340				345					350		
	Pro	Ala	Val	Ala	Ser	Glu	Ala	Ala	Gly	Arg	Leu	Leu	Pro	Ala	Phe	Arg
		355						360					365			
	Glu	Ala	Val	Ala	Arg	Trp	His	Pro	Thr	Ala	Thr	Thr	Ile	Gln	Leu	Leu
20		370					375						380			
	Asp	Pro	Pro	Ala	Ala	Val	Gly	Pro	Val	Trp	Thr	Ala	Arg	Phe	Cys	Phe
	385					390				395						400
	Ser	Gly	Leu	Gln	Ala	Gln	Leu	Leu	Ala	Ala	Gly	Leu	Gly	Glu	Ala	Gly
				405					410							415
25	Leu	Pro	Glu	Arg	Arg	Ala	Gly	Leu	Glu	Arg	Leu	Asp	Ala	Leu	Val	Ala
				420					425					430		
	Ala	Ala	Pro	Ser	Glu	Pro	Trp	Ala	Arg	Ala	Val	Leu	Glu	Arg	Leu	Val
		435					440						445			
	Pro	Asp	Ala	Cys	Asp	Ala	Cys	Pro	Ala	Leu	Arg	Gln	Leu	Leu	Gly	Gly
30		450					455					460				
	Val	Met	Ala	Ala	Val	Cys	Leu	Gln	Ile	Glu	Gln	Thr	Ala	Ser	Ser	Val
	465					470					475					480
	Lys	Phe	Ala	Val	Cys	Gly	Gly	Thr	Gly	Ala	Ala	Phe	Trp	Gly	Leu	Phe
				485					490							495
35	Asn	Val	Asp	Pro	Gly	Asp	Ala	Asp	Ala	Ala	His	Gly	Ala	Ile	Gln	Asp
				500					505					510		
	Ala	Arg	Arg	Ala	Leu	Glu	Ala	Ser	Val	Arg	Ala	Val	Leu	Ser	Ala	Asn
		515							520					525		
	Gly	Ile	Arg	Pro	Arg	Leu	Ala	Pro	Ser	Leu	Ala	Leu	Glu	Gly	Val	Tyr
40		530					535						540			
	Thr	His	Val	Val	Thr	Trp	Ser	Gln	Thr	Gly	Ala	Trp	Phe	Trp	Asn	Ser
	545					550				555						560
	Arg	Asp	Asp	Thr	Asp	Phe	Leu	Gln	Gly	Phe	Pro	Leu	Arg	Gly	Pro	Ala

```

                    565                    570                    575
Tyr Ala Ala Ala Ala Glu Val Met Arg Asp Ala Leu Arg Arg Ile Leu
                    580                    585                    590
Arg Arg Pro Ala Ala Gly Pro Pro Glu Glu Ala Val Cys Ala Arg Ile
5          595                    600                    605
Met Glu Asp Ala Cys Asp Arg Phe Val Leu Asp Ala Phe Gly Arg Arg
        610                    615                    620
Leu Asp Ala Glu Tyr Trp Ser Val Leu Thr Pro Pro Gly Glu Ala Asp
625                    630                    635                    640
10  Asp Pro Leu Pro Gln Thr Ala Phe Arg Gly Gly Ala Leu Leu Asp Ala
        645                    650                    655
Glu Gln Tyr Trp Arg Arg Val Val Arg Val Cys Pro Gly Gly Gly Glu
        660                    665                    670
Ser Val Gly Val Pro Val Asp Leu Tyr Pro Arg Pro Leu Val Leu Pro
15          675                    680                    685
Pro Val Asp Cys Ala His His Leu Arg Glu Ile Leu Arg Glu Ile Gln
        690                    695                    700
Leu Val Phe Thr Gly Val Leu Glu Gly Val Trp Gly Glu Gly Gly Ser
705                    710                    715                    720
20  Phe Val Tyr Pro Phe Glu Glu Lys Met Arg Phe Leu Phe Pro
        725                    730

```

## (2) INFORMATION FOR SEQ ID NO:225:

- 25 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 461 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

30

- (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

```

35  Met Gly Arg Arg Ala Pro Arg Gly Ser Pro Glu Ala Ala Pro Gly Ala
    1          5          10          15
    Asp Val Ala Pro Gly Ala Arg Ala Ala Trp Trp Val Trp Cys Val Gln
          20          25          30
    Val Ala Thr Phe Ile Val Ser Ala Ile Cys Val Val Gly Leu Leu Val
40          35          40          45
    Leu Ala Ser Val Phe Arg Asp Arg Phe Pro Cys Leu Tyr Ala Pro Ala
    50          55          60
    Thr Ser Tyr Ala Glu Ala Asn Ala Thr Val Glu Val Arg Gly Gly Val

```

	65		70		75		80
	Ala Val Pro Leu Arg Leu Asp Thr Gln Ser Leu Leu Ala Thr Tyr Ala						
		85			90		95
	Ile Thr Ser Thr Leu Leu Leu Ala Ala Val Tyr Ala Ala Val Gly						
5		100		105		110	
	Ala Val Thr Ser Arg Tyr Glu Arg Ala Leu Asp Ala Ala Arg Arg Leu						
		115		120		125	
	Ala Ala Ala Arg Met Ala Met Pro His Ala Thr Leu Ile Ala Gly Asn						
		130		135		140	
10	Val Cys Ala Trp Leu Leu Gln Ile Thr Val Leu Leu Leu Ala His Arg						
		145		150		155	160
	Ile Ser Gln Leu Ala His Leu Ile Tyr Val Leu His Phe Ala Cys Leu						
		165		170		175	
	Val Tyr Leu Ala Ala His Phe Cys Thr Arg Gly Val Leu Ser Gly Thr						
15		180		185		190	
	Tyr Leu Arg Gln Val His Gly Leu Ile Asp Pro Ala Pro Thr His His						
		195		200		205	
	Arg Ile Val Gly Pro Val Arg Ala Val Met Thr Asn Ala Leu Leu Leu						
		210		215		220	
20	Gly Thr Leu Leu Cys Thr Ala Ala Ala Ala Val Ser Leu Asn Thr Ile						
		225		230		235	240
	Ala Ala Leu Asn Phe Asn Phe Ser Ala Pro Ser Met Leu Ile Cys Leu						
		245		250		255	
	Thr Thr Leu Phe Ala Leu Leu Val Val Ser Leu Leu Leu Val Val Glu						
25		260		265		270	
	Gly Val Leu Cys His Tyr Val Arg Val Leu Val Gly Pro His Leu Gly						
		275		280		285	
	Ala Ile Ala Ala Thr Gly Ile Val Gly Leu Ala Cys Glu His Tyr His						
		290		295		300	
30	Thr Gly Gly Tyr Tyr Val Val Glu Gln Gln Trp Pro Gly Ala Gln Thr						
		305		310		315	320
	Gly Val Arg Val Val Ala Ala Phe Ala Met Ala Val Leu Arg Cys Thr						
		325		330		335	
	Arg Ala Tyr Leu Tyr His Arg Arg His His Thr Lys Phe Phe Val Arg						
35		340		345		350	
	Met Arg Asp Thr Arg His Arg Ala His Ser Ala Leu Arg Arg Val Arg						
		355		360		365	
	Ser Ser Met Arg Gly Ser Arg Arg Gly Gly Pro Pro Gly Asp Pro Gly						
		370		375		380	
40	Tyr Ala Glu Thr Pro Tyr Ala Ser Val Ser His His Ala Glu Ile Asp						
		385		390		395	400
	Arg Tyr Gly Asp Ser Asp Gly Asp Pro Ile Tyr Asp Glu Val Ala Pro						
		405		410		415	

Asp His Glu Ala Glu Leu Tyr Ala Arg Val Gln Arg Pro Gly Pro Val  
 420 425 430  
 Pro Asp Ala Glu Pro Ile Tyr Asp Thr Val Glu Gly Tyr Ala Pro Arg  
 435 440 445  
 5 Ser Ala Gly Glu Pro Val Tyr Ser Thr Val Arg Arg Trp  
 450 455 460

## (2) INFORMATION FOR SEQ ID NO:226:

## 10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 96 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

20 Met Gly Leu Ala Phe Ser Gly Ala Arg Pro Cys Cys Cys Arg His Asn  
 1 5 10 15  
 Val Ile Ile Thr Asp Gly Gly Glu Val Val Ser Leu Thr Ala His Glu  
 20 25 30  
 Phe Asp Val Val Asp Ile Glu Ser Glu Glu Glu Gly Asn Phe Tyr Val  
 25 35 40 45  
 Pro Pro Asp Met Arg Val Val Thr Arg Ala Pro Gly Pro Gln Tyr Arg  
 50 55 60  
 Arg Ala Ser Asp Pro Pro Ser Arg His Thr Arg Arg Arg Asp Pro Asp  
 65 70 75 80  
 30 Val Ala Arg Pro Pro Ala Thr Leu Thr Pro Pro Leu Ser Asp Ser Glu  
 85 90 95

## (2) INFORMATION FOR SEQ ID NO:227:

## 35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 618 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

Met Ala Ala Ala Ala Thr Pro Gly Ala Lys Arg Pro Ala Asp Pro Ala  
 1 5 10 15  
 Arg Asp Pro Asp Ser Pro Pro Lys Arg Pro Arg Pro Asn Ser Leu Asp  
 5 20 25 30  
 Leu Ala Thr Val Phe Gly Pro Arg Pro Ala Pro Pro Arg Pro Thr Ser  
 35 40 45  
 Pro Gly Ala Pro Gly Ser His Trp Pro Gln Ser Pro Pro Arg Gly Gln  
 50 55 60  
 10 Pro Asp Gly Gly Ala Pro Gly Glu Lys Ala Arg Pro Asp Ala Leu Ser  
 65 70 75 80  
 Glu Ala Ser Ser Gly Pro Pro Thr Pro Asp Ile Pro Leu Ser Pro Gly  
 85 90 95  
 Gly Ala His Ala Ile Asp Pro Asp Cys Ser Pro Gly Pro Pro Asp Pro  
 15 100 105 110  
 Asp Pro Met Trp Ser Ala Ser Ala Ile Pro Asn Ala Leu Pro Pro His  
 115 120 125  
 Ile Leu Ala Glu Thr Phe Glu Arg His Leu Arg Gly Leu Leu Arg Gly  
 130 135 140  
 20 Val Arg Ser Pro Leu Ala Ile Gly Pro Leu Trp Ala Arg Leu Asp Tyr  
 145 150 155 160  
 Leu Cys Ser Leu Val Val Ser Leu Glu Ala Ala Gly Met Val Asp Arg  
 165 170 175  
 Gly Leu Gly Arg His Leu Trp Arg Leu Thr Arg Arg Ala Pro Pro Ser  
 25 180 185 190  
 Ala Ala Glu Ala Val Ala Pro Arg Pro Leu Met Gly Phe Tyr Glu Ala  
 195 200 205  
 Ala Thr Gln Asn Gln Ala Asp Cys Gln Leu Trp Ala Leu Leu Arg Arg  
 210 215 220  
 30 Gly Leu Thr Thr Ala Ser Thr Leu Arg Trp Gly Ala Gln Gly Pro Cys  
 225 230 235 240  
 Phe Ser Ser Gln Trp Leu Thr His Asn Ala Ser Leu Arg Leu Asp Ala  
 245 250 255  
 Gln Ser Ser Ala Val Met Phe Gly Arg Val Asn Glu Pro Thr Ala Arg  
 35 260 265 270  
 Asn Leu Leu Phe Arg Tyr Cys Val Gly Arg Ala Asp Ala Gly Val Asn  
 275 280 285  
 Asp Asp Ala Asp Ala Gly Arg Phe Val Phe His Gln Pro Gly Asp Leu  
 290 295 300  
 40 Ala Glu Glu Asn Val His Ala Cys Gly Val Leu Met Asp Gly His Thr  
 305 310 315 320  
 Gly Met Val Gly Ala Ser Leu Asp Ile Leu Val Cys Pro Arg Asp Pro  
 325 330 335



His Gly Tyr Leu Ala Pro Ala Pro Gln Thr Pro Leu Ala Phe Tyr Glu  
 340 345 350  
 Val Lys Cys Arg Ala Lys Tyr Ala Phe Asp Pro Ala Asp Pro Gly Ala  
 355 360 365  
 5 Pro Ala Ala Ser Ala Tyr Glu Asp Leu Met Ala Arg Arg Ser Pro Glu  
 370 375 380  
 Ala Phe Arg Ala Phe Ile Arg Ser Ile Pro Asn Pro Gly Val Arg Tyr  
 385 390 395 400  
 Phe Ala Pro Gly Arg Val Pro Gly Pro Glu Glu Ala Leu Val Thr Gln  
 10 405 410 415  
 Asp Arg Asp Trp Leu Asp Ser Arg Ala Ala Gly Glu Lys Arg Arg Cys  
 420 425 430  
 Ser Ala Pro Asp Arg Ala Leu Val Glu Leu Asn Ser Gly Val Val Ser  
 435 440 445  
 15 Glu Val Leu Leu Phe Gly Val Pro Asp Leu Glu Arg Arg Thr Ile Ser  
 450 455 460  
 Pro Val Ala Trp Ser Ser Gly Glu Leu Val Arg Arg Glu Pro Ile Phe  
 465 470 475 480  
 Ala Asn Pro Arg His Pro Asn Phe Lys Gln Ile Leu Val Gln Gly Tyr  
 20 485 490 495  
 Val Leu Asp Ser His Phe Pro Asp Cys Pro Leu Gln Pro His Leu Val  
 500 505 510  
 Thr Phe Leu Gly Arg His Arg Ala Gly Ala Glu Glu Gly Val Thr Phe  
 515 520 525  
 25 Arg Leu Glu Asp Gly Arg Gly Ala Pro Ala Gly Arg Gly Gly Ala Pro  
 530 535 540  
 Gly Pro Ala Lys Ala Ser Ile Leu Pro Asp Gln Ala Val Pro Ile Ala  
 545 550 555 560  
 Leu Ile Ile Thr Pro Val Arg Val Glu Pro Gly Ile Tyr Arg Asp Ile  
 30 565 570 575  
 Arg Arg Asn Ser Arg Leu Ala Phe Asp Asp Thr Leu Ala Lys Leu Trp  
 580 585 590  
 Ala Ser Arg Ser Pro Gly Arg Gly Pro Ala Ala Ala Asp Thr Thr Ser  
 595 600 605  
 35 Ser Ser Pro Thr Ala Gly Arg Ser Ser Arg  
 610 615

## (2) INFORMATION FOR SEQ ID NO:228:

- 40 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 516 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

```

Met Asp Glu Ser Gly Arg Gln Arg Pro Ala Ser His Val Ala Ala Asp
 1           5           10           15
Ile Ser Pro Gln Gly Ala His Arg Arg Ser Phe Lys Ala Trp Leu Ala
10           20           25           30
Ser Tyr Ile His Ser Leu Ser Arg Arg Ala Ser Gly Arg Pro Ser Gly
 35           40           45
Pro Ser Pro Arg Asp Gly Ala Val Ser Gly Ala Arg Pro Gly Ser Arg
 50           55           60
15 Arg Arg Ser Ser Phe Arg Glu Arg Leu Arg Ala Gly Leu Ser Arg Trp
 65           70           75           80
Arg Val Ser Arg Ser Ser Arg Arg Arg Ser Ser Pro Glu Ala Pro Gly
 85           90           95
Pro Ala Ala Lys Leu Arg Arg Pro Pro Leu Arg Arg Ser Glu Thr Ala
20           100          105          110
Met Thr Ser Pro Pro Ser Pro Pro Ser His Ile Leu Ser Leu Ala Arg
 115          120          125
Ile His Lys Leu Cys Ile Pro Val Phe Ala Val Asn Pro Ala Leu Arg
 130          135          140
25 Tyr Thr Thr Leu Glu Ile Pro Gly Ala Arg Ser Phe Gly Gly Ser Gly
 145          150          155          160
Gly Tyr Gly Glu Val Gln Leu Ile Arg Glu His Lys Leu Ala Val Lys
 165          170          175
Thr Ile Arg Glu Lys Glu Trp Phe Ala Val Glu Leu Val Ala Thr Leu
30           180          185          190
Leu Val Gly Glu Cys Ala Leu Arg Gly Gly Arg Thr His Asp Ile Arg
 195          200          205
Gly Phe Ile Thr Pro Leu Gly Phe Ser Leu Gln Gln Arg Gln Ile Val
 210          215          220
35 Phe Pro Ala Tyr Asp Met Asp Leu Gly Lys Tyr Ile Gly Gln Leu Ala
 225          230          235          240
Ser Leu Arg Ala Thr Thr Pro Ser Val Ala Thr Ala Leu His His Cys
 245          250          255
Phe Thr Asp Leu Ala Arg Ala Val Val Phe Leu Asn Thr Arg Cys Gly
40           260          265          270
Ile Ser His Leu Asp Ile Lys Cys Ala Asn Val Leu Val Met Leu Arg
 275          280          285
Ser Asp Ala Val Ser Leu Arg Arg Ala Val Leu Ala Asp Phe Ser Leu

```

290                      295                      300  
 Val Thr Leu Asn Ser Asn Ser Thr Ile Ser Arg Gly Gln Phe Cys Leu  
 305                      310                      315                      320  
 Gln Glu Pro Asp Leu Glu Ser Pro Arg Gly Phe Gly Met Pro Ala Ala  
 5                      325                      330                      335  
 Leu Thr Thr Ala Asn Phe His Thr Leu Val Gly His Gly Tyr Asn Gln  
 340                      345                      350  
 Pro Pro Glu Leu Leu Val Lys Tyr Leu Asn Asn Glu Arg Ala Glu Phe  
 355                      360                      365  
 10 Asn Asn Arg Pro Leu Lys His Asp Val Gly Leu Ala Val Asp Leu Tyr  
 370                      375                      380  
 Ala Leu Gly Gln Thr Leu Leu Glu Leu Leu Val Ser Val Tyr Val Ala  
 385                      390                      395                      400  
 Pro Ser Leu Gly Val Pro Val Thr Arg Val Pro Gly Tyr Gln Tyr Phe  
 15                      405                      410                      415  
 Asn Asn Gln Leu Ser Pro Asp Phe Ala Val Leu Ala Tyr Arg Cys Val  
 420                      425                      430  
 Leu His Pro Ala Leu Phe Val Asn Ser Ala Glu Thr Asn Thr His Gly  
 435                      440                      445  
 20 Leu Ala Tyr Asp Val Pro Glu Gly Ile Arg Arg His Leu Arg Asn Pro  
 450                      455                      460  
 Lys Ile Arg Arg Ala Phe Thr Glu Gln Cys Ile Asn Tyr Gln Arg Thr  
 465                      470                      475                      480  
 His Lys Ala Val Leu Ser Ser Val Ser Leu Pro Pro Glu Leu Arg Pro  
 25                      485                      490                      495  
 Leu Leu Val Leu Val Ser Arg Leu Cys His Ala Asn Pro Ala Ala Arg  
 500                      505                      510  
 His Ser Leu Ser  
 515  
 30

## (2) INFORMATION FOR SEQ ID NO:229:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 217 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

Met Ser Arg Asp Ala Ser His Ala Ala Leu Arg Arg Arg Leu Ala Glu

1                      5                      10                      15  
 Thr His Leu Arg Ala Glu Val Tyr Arg Asp Gln Thr Leu Gln Leu His  
                          20                      25                      30  
 Arg Glu Gly Val Ser Thr Gln Asp Pro Arg Phe Val Gly Ala Phe Met  
 5                      35                      40                      45  
 Ala Ala Lys Ala Ala His Leu Glu Leu Glu Ala Arg Leu Lys Ser Arg  
                          50                      55                      60  
 Ala Arg Leu Glu Met Met Arg Gln Arg Ala Thr Cys Val Lys Ile Arg  
 65                      70                      75                      80  
 10 Val Glu Glu Gln Ala Ala Arg Arg Asp Phe Leu Thr Ala His Arg Arg  
                          85                      90                      95  
 Tyr Leu Asp Pro Ala Leu Ser Leu Asp Ala Ala Asp Asp Arg Leu Ala  
                          100                      105                      110  
 Asp Gln Glu Glu Gln Leu Glu Glu Ala Ala Ala Asn Ala Ser Leu Trp  
 15                      115                      120                      125  
 Gly Asp Gly Asp Leu Ala Asp Gly Trp Met Ser Pro Gly Asp Ser Asp  
                          130                      135                      140  
 Leu Leu Val Met Trp Gln Leu Thr Ser Ala Pro Lys Val His Thr Asp  
 145                      150                      155                      160  
 20 Ala Pro Ser Arg Pro Gly Ser Arg Pro Thr Tyr Thr Pro Ser Ala Ala  
                          165                      170                      175  
 Gly Arg Pro Asp Ala Gln Ala Ala Pro Pro Pro Glu Thr Ala Pro Ser  
                          180                      185                      190  
 Pro Glu Pro Ala Pro Gly Pro Ala Ala Asp Pro Ala Ser Gly Ser Gly  
 25                      195                      200                      205  
 Phe Ala Arg Asp Cys Pro Asp Gly Glu  
                          210                      215

## (2) INFORMATION FOR SEQ ID NO:230:

30

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 430 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

35

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

40

Met Phe Gly Gln Gln Leu Ala Ser Asp Val Gln Gln Tyr Leu Glu Arg  
 1                      5                      10                      15  
 Leu Glu Lys Gln Arg Gln Gln Lys Val Gly Val Asp Glu Ala Ser Ala

	20		25		30
	Gly Leu Thr	Leu Gly Gly Asp Ala	Leu Arg Val	Pro Phe	Leu Asp Phe
	35		40		45
	Ala Thr Ala Thr	Pro Lys Arg His	Gln Thr Val	Val Pro Gly	Val Gly
5	50		55		60
	Thr Leu His Asp	Cys Cys Glu His	Ser Pro Leu	Phe Ser Ala	Val Ala
	65		70		75
	Arg Arg Leu Leu	Phe Asn Ser Leu	Val Pro Ala	Gln Leu Arg	Gly Arg
		85		90	
10	Asp Phe Gly Gly	Asp His Thr Ala	Lys Leu Glu	Phe Leu Ala	Pro Glu
	100		105		110
	Leu Val Arg Ala	Val Ala Arg Leu	Arg Phe Arg	Glu Cys Ala	Pro Glu
	115		120		125
	Asp Ala Val Pro	Gln Arg Asn Ala	Tyr Tyr Ser	Val Leu Asn	Thr Phe
15	130		135		140
	Gln Ala Leu His	Arg Ser Glu Ala	Phe Arg Gln	Leu Val His	Phe Val
	145		150		155
	Arg Asp Phe Ala	Gln Leu Leu Lys	Thr Ser Phe	Arg Ala Ser	Ser Leu
		165		170	
20	Ala Glu Thr Thr	Gly Pro Pro Lys	Lys Arg Ala	Lys Val Asp	Val Ala
	180		185		190
	Thr His Gly Gln	Thr Tyr Gly Thr	Leu Glu Leu	Phe Gln Lys	Met Ile
	195		200		205
	Leu Met His Ala	Thr Tyr Phe Leu	Ala Ala Val	Leu Leu Gly	Asp His
25	210		215		220
	Ala Glu Gln Val	Asn Thr Phe Leu	Arg Leu Val	Phe Glu Ile	Pro Leu
	225		230		235
	Phe Ser Asp Thr	Ala Val Arg His	Phe Arg Gln	Arg Ala Thr	Val Phe
		245		250	
30	Leu Val Pro Arg	Arg His Gly Lys	Thr Trp Phe	Leu Val Pro	Leu Ile
	260		265		270
	Ala Leu Ser Leu	Ala Ser Phe Arg	Gly Ile Lys	Ile Gly Tyr	Thr Ala
	275		280		285
	His Ile Arg Lys	Ala Thr Glu Pro	Val Phe Asp	Glu Ile Asp	Ala Cys
35	290		295		300
	Leu Arg Gly Trp	Phe Gly Ser Ser	Arg Val Asp	His Val Lys	Gly Glu
	305		310		315
	Thr Ile Ser Phe	Ser Phe Pro Asp	Gly Ser Arg	Ser Thr Ile	Val Phe
		325		330	
40	Ala Ser Ser His	Asn Thr Asn Val	Ser Thr Pro	Ser Ser Arg	Gly Ala
	340		345		350
	Cys Phe Pro Gly	Ala Ala Leu Pro	Glu Ile Asp	Arg Gln Thr	Asn Thr
	355		360		365

Ala Arg Arg Glu Cys Gly Thr Trp Gln Pro Pro Pro Pro Trp Arg Gly  
 370 375 380  
 Glu Ala Leu Leu Phe Ile Cys Asn Arg Thr Met Arg Leu Trp Pro Arg  
 385 390 395 400  
 5 Pro Ala Arg Pro Arg Gly Ser Ser Leu Gln Thr Gly Gly Trp Tyr Thr  
 405 410 415  
 Met Thr Glu Arg Arg Gly Ala Thr Arg Arg Trp Ser Gly Gly  
 420 425 430

10 (2) INFORMATION FOR SEQ ID NO:231:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 315 amino acids  
 (B) TYPE: amino acid  
 15 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:

Val Trp Arg Val Val Arg Gly Asp Glu Arg Leu Lys Ile Phe Arg Cys  
 1 5 10 15  
 Leu Thr Val Leu Thr Glu Pro Leu Cys Gln Val Pro Asp Pro Asp Pro  
 25 20 25 30  
 Glu Arg Ala Leu Phe Cys Glu Ile Phe Leu Tyr Leu Trp Lys Ala Leu  
 35 40 45  
 Arg Leu Pro Ser Asn Thr Phe Phe Ala Ile Phe Phe Phe Asn Arg Glu  
 50 55 60  
 30 Arg Arg Tyr Cys Ala Thr Val His Leu Arg Ser Val Thr His Pro Arg  
 65 70 75 80  
 Thr Pro Leu Leu Cys Thr Leu Ala Phe Gly His Leu Glu Ala Asp Pro  
 85 90 95  
 Glu Glu Thr Pro Asp Pro Ala Ala Glu Gln Leu Ala Asp Glu Pro Val  
 35 100 105 110  
 Ala His Glu Leu Asp Gly Ala Tyr Leu Val Pro Thr Glu Pro Pro Pro  
 115 120 125  
 Asn Pro Gly Ala Cys Cys Ala Leu Gly Pro Gly Ala Trp Trp His Leu  
 130 135 140  
 40 Pro Gly Gly Arg Ile Tyr Cys Trp Ala Met Asp Asp Asp Leu Gly Ser  
 145 150 155 160  
 Leu Cys Pro Pro Gly Ser Arg Ala Arg His Leu Gly Trp Leu Leu Ser  
 165 170 175

Arg Ile Thr Asp Pro Pro Gly Gly Gly Gly Ala Cys Ala Pro Thr Ala  
 180 185 190  
 His Ile Asp Ser Ala Asn Ala Leu Trp Arg Ala Pro Ala Val Ala Glu  
 195 200 205  
 5 Ala Cys Pro Cys Val Ala Pro Cys Met Trp Ser Asn Met Ala Gln Arg  
 210 215 220  
 Thr Leu Ala Val Arg Gly Asp Ala Ser Leu Cys Gln Leu Leu Phe Gly  
 225 230 235 240  
 His Pro Val Asp Ala Val Ile Leu Arg Gln Ala Thr Arg Arg Pro Arg  
 10 245 250 255  
 Ile Thr Ala His Leu His Glu Val Val Val Gly Arg Asp Gly Ala Glu  
 260 265 270  
 Ser Val Ile Arg Pro Thr Ser Ala Gly Trp Arg Leu Cys Val Leu Ser  
 275 280 285  
 15 Ser Tyr Thr Ser Arg Leu Phe Ala Thr Ser Cys Pro Ala Val Ala Arg  
 290 295 300  
 Ala Val Ala Arg Ala Ser Ser Ser Asp Tyr Lys  
 305 310 315

20 (2) INFORMATION FOR SEQ ID NO:232:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 698 amino acids  
 (B) TYPE: amino acid  
 25 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

Met Asn Ala His Phe Ala Asn Glu Val Gln Tyr Asp Leu Thr Arg Asp  
 1 5 10 15  
 Pro Ser Ser Pro Ala Ser Leu Ile His Val Ile Ile Ser Ser Glu Cys  
 20 25 30  
 35 Leu Ala Ala Ala Gly Val Pro Leu Ser Ala Leu Val Arg Gly Arg Pro  
 35 40 45  
 Asp Gly Gly Ala Ala Ala Asn Phe Arg Val Glu Thr Gln Thr Arg Ala  
 50 55 60  
 40 His Ala Thr Gly Asp Cys Thr Pro Trp Arg Ser Ala Phe Ala Ala Tyr  
 65 70 75 80  
 Val Pro Ala Asp Ala Val Gly Ala Ile Leu Ala Pro Val Ile Pro Ala  
 85 90 95

His Pro Asp Leu Leu Pro Arg Val Pro Ser Ala Gly Gly Leu Phe Val  
 100 105 110  
 Ser Leu Pro Val Ala Cys Asp Ala Gln Gly Val Tyr Asp Pro Tyr Thr  
 115 120 125  
 5 Val Ala Ala Leu Arg Leu Ala Trp Gly Pro Trp Ala Thr Cys Ala Arg  
 130 135 140  
 Val Leu Leu Phe Ser Tyr Asp Glu Leu Val Pro Pro Asn Thr Arg Tyr  
 145 150 155 160  
 Ala Ala Asp Gly Ala Arg Leu Met Arg Leu Cys Arg His Phe Cys Arg  
 165 170 175  
 10 Tyr Val Ala Arg Leu Gly Ala Ala Ala Pro Ala Ala Ala Thr Glu Ala  
 180 185 190  
 Ala Ala His Leu Ser Leu Gly Met Gly Glu Ser Gly Thr Pro Thr Pro  
 195 200 205  
 15 Gln Ala Ser Ser Val Ser Gly Gly Ala Gly Pro Ala Val Val Gly Thr  
 210 215 220  
 Pro Asp Pro Pro Ile Ser Pro Glu Glu Gln Leu Thr Ala Pro Gly Gly  
 225 230 235 240  
 Asp Thr Ala Thr Ala Glu Asp Val Ser Ile Thr Gln Glu Asn Glu Glu  
 245 250 255  
 20 Ile Leu Ala Leu Val Gln Arg Ala Val Gln Asp Val Thr Arg Arg His  
 260 265 270  
 Pro Val Arg Ala Arg Pro Lys His Ala Ala Ser Gly Val Ala Ser Gly  
 275 280 285  
 25 Leu Arg Gln Gly Ala Leu Val His Gln Ala Val Ser Gly Gly Ala Leu  
 290 295 300  
 Gly Ala Ser Asp Ala Glu Ala Val Leu Ala Gly Leu Glu Pro Pro Gly  
 305 310 315 320  
 Gly Gly Arg Phe Ala Thr Pro Gly Gly Pro Arg Ala Ala Gly Glu Asp  
 325 330 335  
 30 Val Leu Asn Asp Val Leu Thr Leu Val Pro Gly Thr Ala Lys Pro Arg  
 340 345 350  
 Ser Leu Val Glu Trp Leu Asp Arg Gly Trp Glu Ala Gly Gly Asp Arg  
 355 360 365  
 35 Pro Asp Trp Leu Trp Ser Arg Arg Ser Ile Ser Val Val Leu Arg His  
 370 375 380  
 His Tyr Gly Thr Lys Gln Arg Phe Val Val Val Ser Tyr Glu Asn Ser  
 385 390 395 400  
 Val Ala Trp Gly Gly Arg Arg Ala Arg Pro Pro Arg Leu Ser Ser Glu  
 405 410 415  
 40 Leu Ala Thr Ala Leu Thr Glu Ala Cys Ala Ala Glu Arg Val Val Arg  
 420 425 430  
 Pro His Gln Leu Ser Pro Ala Ala Gln Thr Ala Leu Leu Arg Arg Phe



```

          435              440              445
Pro Ala Leu Glu Gly Pro Leu Arg His Pro Arg Pro Val Leu Gln Pro
          450              455              460
Phe Asp Ile Ala Ala Glu Val Ala Phe Val Ala Arg Ile Gln Ile Ala
5  465              470              475              480
Cys Leu Arg Ala Leu Gly His Ser Ile Arg Ala Ala Leu Gln Gly Gly
          485              490              495
Pro Arg Ile Phe Gln Arg Leu Arg Tyr Asp Phe Gly Pro His Gln Ser
          500              505              510
10 Glu Trp Leu Gly Glu Val Thr Arg Arg Phe Pro Val Leu Leu Glu Asn
          515              520              525
Leu Met Arg Ala Leu Glu Gly Thr Ala Pro Asp Ala Phe Phe His Thr
          530              535              540
Ala Tyr Ala Val Leu Ala His Leu Gly Gly Gln Gly Gly Arg Gly Arg
15 545              550              555              560
Arg Arg Arg Leu Val Pro Leu Ser Asp Asp Ile Pro Ala Arg Phe Ala
          565              570              575
Asp Ser Asp Ala His Tyr Ala Phe Asp Tyr Tyr Ser Thr Ser Gly Asp
          580              585              590
20 Thr Leu Arg Leu Thr Asn Arg Pro Ile Ala Val Val Ile Asp Gly Asp
          595              600              605
Val Asn Gly Arg Glu Gln Ser Lys Cys Arg Phe Met Glu Gly Ser Pro
          610              615              620
Ser Thr Ala Pro His Arg Val Cys Glu Gln Tyr Leu Pro Gly Glu Ser
25 625              630              635              640
Tyr Ala Tyr Leu Cys Leu Gly Phe Asn Arg Arg Leu Cys Gly Leu Val
          645              650              655
Val Phe Pro Gly Gly Phe Ala Phe Thr Ile Asn Thr Ala Ala Tyr Leu
          660              665              670
30 Ser Leu Ala Asp Pro Val Ala Arg Ala Val Gly Leu Arg Phe Cys Arg
          675              680              685
Gly Ala Ala Thr Gly Pro Gly Leu Val Arg
          690              695

```

35 (2) INFORMATION FOR SEQ ID NO:233:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 423 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:

Val Pro Glu Gly Ala Trp Val Gly Gly Ala Cys Ala Arg Pro Arg Gly  
 5     1                     5                     10                     15  
 Pro Arg Ala His Val Arg Leu Tyr Ala Val Cys Phe Val Cys Pro Gln  
                    20                     25                     30  
 Gly Ile Arg Gly Gln Asp Phe Asn Leu Leu Phe Val Asp Glu Ala Asn  
                    35                     40                     45  
 10 Phe Ile Arg Pro Asp Ala Val Gln Thr Ile Met Gly Phe Leu Asn Gln  
                    50                     55                     60  
 Ala Asn Cys Lys Ile Ile Phe Val Ser Ser Thr Asn Thr Gly Lys Ala  
 65                     70                     75                     80  
 Ser Thr Ser Phe Leu Tyr Asn Leu Arg Gly Ala Ala Asp Glu Leu Leu  
 15                     85                     90                     95  
 Asn Val Val Thr Tyr Ile Cys Asp Asp His Met Pro Arg Val Val Thr  
                    100                     105                     110  
 His Thr Asn Ala Thr Ala Cys Ser Cys Tyr Ile Leu Asn Lys Pro Val  
                    115                     120                     125  
 20 Phe Ile Thr Met Asp Gly Ala Val Arg Arg Thr Ala Asp Leu Phe Leu  
                    130                     135                     140  
 Pro Asp Ser Phe Met Gln Glu Ile Ile Gly Gly Gln Ala Arg Glu Thr  
 145                     150                     155                     160  
 Gly Asp Asp Arg Pro Val Leu Thr Lys Ser Ala Gly Glu Arg Phe Leu  
 25                     165                     170                     175  
 Leu Tyr Arg Pro Ser Thr Thr Thr Asn Ser Gly Leu Met Ala Pro Glu  
                    180                     185                     190  
 Leu Tyr Val Tyr Val Asp Pro Ala Phe Thr Ala Asn Thr Arg Ala Ser  
                    195                     200                     205  
 30 Gly Thr Gly Ile Ala Val Val Gly Arg Tyr Arg Asp Asp Phe Ile Ile  
                    210                     215                     220  
 Phe Ala Leu Glu His Phe Phe Leu Arg Ala Leu Thr Gly Ser Ala Pro  
 225                     230                     235                     240  
 Ala Asp Ile Ala Arg Cys Val Val His Ser Leu Ala Gln Val Leu Ala  
 35                     245                     250                     255  
 Leu His Pro Gly Ala Phe Arg Ser Val Arg Val Ala Val Glu Gly Asn  
                    260                     265                     270  
 Ser Ser Gln Asp Ser Ala Val Ala Ile Ala Thr His Val His Thr Glu  
                    275                     280                     285  
 40 Met His Arg Ile Leu Ala Ser Ala Gly Ala Asn Gly Pro Gly Pro Glu  
                    290                     295                     300  
 Leu Leu Phe Tyr His Cys Glu Pro Pro Gly Gly Ala Val Leu Tyr Pro  
 305                     310                     315                     320

[illegible]

(2) INFORMATION FOR SEQ ID NO:234:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

	Met	Ile	Thr	Asp	Cys	Phe	Glu	Ala	Asp	Ile	Ala	Ile	Pro	Ser	Gly	Ile
	1				5					10					15	
30	Ser	Arg	Pro	Asp	Ala	Ala	Ala	Leu	Gln	Arg	Cys	Glu	Gly	Arg	Val	Val
				20					25					30		
	Phe	Leu	Pro	Thr	Ile	Arg	Arg	Gln	Leu	Ala	Asp	Val	Ala	His	Glu	Ser
				35				40					45			
	Phe	Val	Ser	Gly	Gly	Val	Ser	Pro	Asp	Thr	Leu	Gly	Leu	Leu	Leu	Ala
35		50					55				60					
	Tyr	Arg	Arg	Arg	Phe	Pro	Ala	Val	Ile	Thr	Arg	Val	Leu	Pro	Thr	Arg
	65				70					75					80	
	Ile	Val	Ala	Cys	Pro	Val	Asp	Leu	Gly	Leu	Thr	His	Ala	Gly	Thr	Val
				85				90					95			
40	Asn	Leu	Arg	Asn	Thr	Ser	Pro	Val	Asp	Leu	Cys	Asn	Gly	Asp	Pro	Val
				100				105					110			
	Ser	Leu	Val	Pro	Pro	Val	Phe	Glu	Gly	Gln	Ala	Thr	Asp	Val	Arg	Leu
				115				120					125			

Glu Ser Leu Asp Leu Thr Leu Arg Phe Pro Val Pro Leu Pro Thr Pro  
 130 135 140  
 Leu Ala Arg Glu Ile Val Ala Arg Leu Val Arg Ile Arg Asp Leu Asn  
 145 150 155 160  
 5 Pro Asp Pro Arg Thr Pro Gly Glu Leu Pro Asp Leu Asn Val Leu Tyr  
 165 170 175  
 Tyr Asn Gly Ala Arg Leu Ser Leu Val Ala Asp Val Gln Gln Leu Ala  
 180 185 190  
 Ser Val Asn Thr Glu Leu Arg Ser Leu Val Leu Asn Met Val Tyr Ser  
 10 195 200 205  
 Ile Thr Glu Gly Thr Thr Leu Ile Leu Thr Leu Ile Pro Arg Leu Leu  
 210 215 220  
 Ala Leu Ser Ala Gln Asp Gly Tyr Val Asn Ala Leu Leu Gln Met Gln  
 225 230 235 240  
 15 Ser Val Thr Arg Glu Ala Ala Gln Leu Ile His Pro Glu Ala Pro Met  
 245 250 255  
 Leu Met Gln Asp Gly Glu Arg Arg Leu Pro Leu Tyr Glu Ala Leu Val  
 260 265 270  
 Ala Trp Leu Ala His Ala Gly Gln Leu Gly Asp Ile Leu Ala Pro Ala  
 20 275 280 285  
 Val Arg Val Cys Thr Phe Asp Gly Ala Ala Val Val Gln Ser Gly Asp  
 290 295 300  
 Met Ala Pro Val Ile Arg Tyr Pro  
 305 310

25

## (2) INFORMATION FOR SEQ ID NO:235:

## (i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 222 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

Met Thr Met Arg Asp Asp Val Pro Leu Leu Asp Arg Glu Leu Val Tyr  
 1 5 10 15  
 40 Glu Ala Ala Cys Gly Gly Glu Asp Gly Glu Leu Pro Leu Asp Glu Gln  
 20 25 30  
 Phe Ser Leu Ser Ser Tyr Gly Thr Ser Asp Phe Phe Val Ser Ser Ala  
 35 40 45

Tyr Ser Arg Leu Pro Pro His Thr Gln Pro Val Phe Ser Lys Arg Val  
 50 55 60  
 Val Met Phe Ala Trp Ser Phe Leu Val Leu Lys Pro Leu Glu Leu Val  
 65 70 75 80  
 5 Ala Ala Gly Met Tyr Tyr Gly Trp Thr Gly Arg Ala Val Ala Pro Ala  
 85 90 95  
 Cys Ile Ile Ala Ala Val Leu Ala Tyr Tyr Val Thr Trp Leu Ala Arg  
 100 105 110  
 Ala Leu Leu Leu Tyr Val Asn Ile Lys Arg Asp Arg Leu Pro Leu Ser  
 10 115 120 125  
 Pro Pro Val Phe Trp Gly Leu Cys Val Ile Met Gly Gly Ala Ala Leu  
 130 135 140  
 Cys Ala Leu Val Ala Ala His Glu Thr Phe Ser Pro Asp Gly Leu  
 145 150 155 160  
 15 Phe His Trp Ile Thr Ala Ser Gln Leu Leu Pro Arg Thr Asp Pro Leu  
 165 170 175  
 Arg Ala Arg Ser Leu Gly Ile Ala Cys Ala Ala Gly Ala Ala Met Trp  
 180 185 190  
 Val Ala Ala Asp Cys Phe Ala Ala Phe Thr Asn Phe Phe Leu Ala  
 20 195 200 205  
 Arg Phe Trp Thr Arg Ala Ile Leu Lys Ala Pro Val Ala Phe  
 210 215 220

## (2) INFORMATION FOR SEQ ID NO:236:

25

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 824 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 30 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:236:

35

Met Gly Pro Gly Leu Trp Val Val Met Gly Val Leu Val Gly Val Ala  
 1 5 10 15  
 Gly Gly His Asp Thr Tyr Trp Thr Glu Gln Ile Asp Pro Trp Phe Leu  
 20 25 30  
 40 His Gly Leu Gly Leu Ala Arg Thr Tyr Trp Arg Asp Thr Asn Thr Gly  
 35 40 45  
 Arg Leu Trp Leu Pro Asn Thr Pro Asp Ala Ser Asp Pro Gln Arg Gly  
 50 55 60

Arg Leu Ala Pro Pro Gly Glu Leu Asn Leu Thr Thr Ala Ser Val Pro  
 65 70 75 80  
 Met Leu Arg Trp Tyr Ala Glu Arg Phe Cys Phe Val Leu Val Thr Thr  
 85 90 95  
 5 Ala Glu Phe Pro Arg Asp Pro Gly Gln Leu Leu Tyr Ile Pro Lys Thr  
 100 105 110  
 Tyr Leu Leu Gly Arg Pro Arg Asn Ala Ser Leu Pro Glu Leu Pro Glu  
 115 120 125  
 Ala Gly Pro Thr Ser Arg Pro Pro Ala Glu Val Thr Gln Leu Lys Gly  
 10 130 135 140  
 Leu Ser His Asn Pro Gly Ala Ser Ala Leu Leu Arg Ser Arg Ala Trp  
 145 150 155 160  
 Val Thr Phe Ala Ala Ala Pro Asp Arg Glu Gly Leu Thr Phe Pro Arg  
 165 170 175  
 15 Gly Asp Asp Gly Ala Thr Glu Arg His Pro Asp Gly Arg Arg Asn Ala  
 180 185 190  
 Pro Pro Pro Gly Pro Pro Ala Gly Thr Pro Arg His Pro Thr Thr Asn  
 195 200 205  
 Leu Ser Ile Ala His Leu His Asn Ala Ser Val Thr Trp Leu Ala Arg  
 20 210 215 220  
 Leu Leu Arg Thr Pro Gly Arg Tyr Val Tyr Leu Ser Pro Ser Ala Ser  
 225 230 235 240  
 Thr Trp Pro Val Gly Val Trp Thr Thr Gly Gly Leu Ala Phe Gly Cys  
 245 250 255  
 25 Asp Ala Ala Leu Val Arg Ala Arg Tyr Gly Lys Gly Phe Met Gly Leu  
 260 265 270  
 Val Ile Ser Met Arg Asp Ser Pro Pro Ala Glu Ile Ile Val Val Pro  
 275 280 285  
 Ala Asp Lys Thr Leu Ala Arg Val Gly Asn Pro Thr Asp Glu Asn Ala  
 30 290 295 300  
 Pro Ala Val Leu Pro Gly Pro Pro Ala Gly Pro Arg Tyr Arg Val Phe  
 305 310 315 320  
 Val Leu Gly Ala Pro Thr Pro Ala Asp Asn Gly Ser Ala Leu Asp Ala  
 325 330 335  
 35 Leu Arg Arg Val Ala Gly Tyr Pro Glu Glu Ser Thr Asn Tyr Ala Gln  
 340 345 350  
 Tyr Met Ser Arg Ala Tyr Ala Glu Phe Leu Gly Glu Asp Pro Gly Ser  
 355 360 365  
 Gly Thr Asp Ala Arg Pro Ser Leu Phe Trp Arg Leu Ala Gly Leu Leu  
 40 370 375 380  
 Ala Ser Ser Gly Phe Ala Phe Val Asn Ala Ala His Ala His Asp Ala  
 385 390 395 400  
 Ile Arg Leu Ser Asp Leu Leu Gly Phe Leu Ala His Ser Arg Val Leu

405 410 415  
 Ala Gly Leu Ala Arg Ala Ala Gly Cys Ala Ala Asp Ser Val Phe Leu  
 420 425 430  
 Asn Val Ser Val Leu Asp Pro Ala Ala Arg Leu Arg Leu Glu Ala Arg  
 5 435 440 445  
 Leu Gly His Leu Val Ala Ala Ile Arg Glu Gln Ser Leu Ala Ala His  
 450 455 460  
 Ala Leu Gly Tyr Gln Leu Ala Phe Val Leu Asp Ser Pro Ala Ala Tyr  
 465 470 475 480  
 10 Gly Ala Val Ala Pro Ser Ala Ala Arg Leu Ile Asp Ala Leu Tyr Ala  
 485 490 495  
 Glu Phe Leu Gly Gly Arg Ala Leu Thr Ala Pro Met Val Arg Arg Ala  
 500 505 510  
 Leu Phe Tyr Ala Thr Ala Val Leu Arg Ala Pro Phe Leu Ala Gly Ala  
 15 515 520 525  
 Pro Ser Ala Glu Gln Arg Glu Arg Ala Arg Arg Gly Leu Leu Ile Thr  
 530 535 540  
 Thr Ala Leu Cys Thr Ser Asp Val Ala Ala Ala Thr His Ala Asp Leu  
 545 550 555 560  
 20 Arg Ala Ala Arg Thr Asp His Gln Lys Asn Leu Phe Trp Leu Pro Asp  
 565 570 575  
 His Phe Ser Pro Cys Ala Ala Ser Leu Arg Phe Asp Leu Ala Glu Gly  
 580 585 590  
 Gly Phe Ile Leu Asp Ala Met Ala Thr Arg Ser Asp Ile Pro Ala Asp  
 25 595 600 605  
 Val Met Ala Gln Gln Thr Arg Gly Val Ala Ser Val Leu Thr Arg Trp  
 610 615 620  
 Ala His Tyr Asn Ala Leu Ile Arg Ala Phe Val Pro Glu Ala Thr His  
 625 630 635 640  
 30 Gln Cys Ser Gly Pro Ser His Asn Ala Glu Pro Arg Ile Leu Val Pro  
 645 650 655  
 Ile Thr His Asn Ala Ser Tyr Val Val Thr His Thr Pro Leu Pro Arg  
 660 665 670  
 Gly Ile Gly Tyr Lys Leu Thr Gly Val Asp Val Arg Arg Pro Leu Phe  
 35 675 680 685  
 Ile Thr Tyr Leu Thr Ala Thr Cys Glu Gly His Ala Arg Glu Ile Glu  
 690 695 700  
 Pro Lys Arg Leu Val Arg Thr Glu Asn Arg Arg Asp Leu Gly Leu Val  
 705 710 715 720  
 40 Gly Ala Val Phe Leu Arg Tyr Thr Pro Ala Gly Glu Val Met Ser Val  
 725 730 735  
 Leu Leu Val Asp Thr Asp Ala Thr Gln Gln Leu Ala Gln Gly Pro  
 740 745 750  
 600

Val Ala Gly Thr Pro Asn Val Phe Ser Ser Asp Val Pro Ser Val Leu  
 755 760 765  
 Leu Phe Pro Asn Gly Thr Val Ile His Leu Leu Ala Phe Asp Thr Leu  
 770 775 780  
 5 Pro Ile Ala Thr Ile Ala Pro Gly Phe Leu Ala Ala Ser Ala Leu Gly  
 785 790 795 800  
 Val Val Met Ile Thr Ala Ala Gly Ile Leu Arg Val Val Arg Thr Cys  
 805 810 815  
 Val Pro Phe Leu Trp Arg Arg Glu  
 10 820

## (2) INFORMATION FOR SEQ ID NO:237:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 370 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

Met Ala Ser His Ala Gly Gln Gln His Ala Pro Ala Phe Gly Gln Ala  
 25 1 5 10 15  
 Ala Arg Ala Ser Gly Pro Thr Asp Gly Arg Ala Ala Ser Arg Pro Ser  
 20 25 30  
 His Arg Gln Gly Ala Ser Asp Pro Glu Leu Pro Thr Leu Leu Arg Val  
 35 40 45  
 30 Tyr Ile Asp Gly Pro His Gly Val Gly Lys Thr Thr Thr Ser Ala Gln  
 50 55 60  
 Leu Met Glu Ala Leu Gly Pro Arg Asp Asn Ile Val Tyr Val Pro Glu  
 65 70 75 80  
 Pro Met Thr Tyr Trp Gln Val Leu Gly Ala Ser Glu Thr Leu Thr Asn  
 35 85 90 95  
 Ile Tyr Asn Thr Gln His Arg Leu Asp Arg Gly Glu Ile Ser Ala Gly  
 100 105 110  
 Glu Ala Ala Val Val Met Thr Ser Ala Gln Ile Thr Met Ser Thr Pro  
 115 120 125  
 40 Tyr Ala Ala Thr Asp Ala Val Leu Ala Pro His Ile Gly Gly Glu Ala  
 130 135 140  
 Val Gly Pro Gln Ala Pro Pro Ala Leu Thr Leu Val Phe Asp Arg  
 145 150 155 160



His Pro Ile Ala Ser Leu Leu Cys Tyr Pro Ala Ala Arg Tyr Leu Met.  
                                   165                                  170                                  175  
 Gly Ser Met Thr Pro Gln Ala Val Leu Ala Phe Val Met Pro Pro Thr  
                                   180                                  185                                  190  
 5 Ala Pro Gly Thr Asn Leu Val Leu Gly Val Leu Pro Glu Ala Glu His  
                                   195                                  200                                  205  
 Ala Asp Arg Leu Ala Arg Arg Gln Arg Pro Gly Glu Arg Leu Asp Leu  
                                   210                                  215                                  220  
 Ala Met Leu Ser Ala Ile Arg Arg Val Tyr Asp Leu Leu Ala Asn Thr  
 10 225                                  230                                  235                                  240  
 Val Arg Tyr Leu Gln Arg Gly Gly Arg Trp Arg Glu Asp Trp Gly Arg  
                                   245                                  250                                  255  
 Leu Thr Gly Val Ala Ala Ala Thr Pro Arg Pro Asp Pro Glu Asp Gly  
                                   260                                  265                                  270  
 15 Ala Gly Ser Leu Pro Arg Ile Glu Asp Thr Leu Phe Ala Leu Phe Arg  
                                   275                                  280                                  285  
 Val Pro Glu Leu Leu Ala Pro Asn Gly Asp Leu Tyr His Ile Phe Ala  
                                   290                                  295                                  300  
 Trp Val Leu Asp Val Leu Ala Asp Arg Leu Leu Pro Met His Leu Phe  
 20 305                                  310                                  315                                  320  
 Val Leu Asp Tyr Asp Gln Ser Pro Val Gly Cys Arg Asp Ala Leu Leu  
                                   325                                  330                                  335  
 Arg Leu Thr Ala Gly Met Ile Pro Thr Arg Val Thr Thr Ala Gly Ser  
                                   340                                  345                                  350  
 25 Ile Ala Glu Ile Arg Asp Leu Ala Arg Thr Phe Ala Arg Glu Val Gly  
                                   355                                  360                                  365  
 Gly Val  
                                   370

30 (2) INFORMATION FOR SEQ ID NO:238:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 279 amino acids  
 (B) TYPE: amino acid  
 35 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

Met Ala Arg Thr Gly Arg Arg Ala Ala Val Gly Arg Pro Ala Arg Thr  
 1                                  5                                  10                                  15

Ser Ser Leu Thr Glu Arg Arg Arg Val Leu Leu Ala Gly Val Arg Ser  
 20 25 30  
 His Thr Arg Phe Tyr Lys Ala Phe Ala Arg Glu Val Arg Glu Phe Asn  
 35 40 45  
 5 Ala Thr Arg Ile Cys Gly Thr Leu Leu Thr Leu Met Ser Gly Ser Leu  
 50 55 60  
 Gln Gly Arg Ser Leu Phe Glu Ala Thr Arg Val Thr Leu Ile Cys Glu  
 65 70 75 80  
 Val Asp Leu Gly Pro Arg Arg Pro Asp Cys Ile Cys Val Phe Glu Phe  
 85 90 95  
 10 Ala Asn Asp Lys Thr Leu Gly Gly Val Cys Val Ile Leu Lys Thr Cys  
 100 105 110  
 Lys Ser Ile Ser Ser Gly Asp Thr Ala Ser Lys Arg Glu Gln Arg Thr  
 115 120 125  
 15 Thr Gly Met Lys Gln Leu Arg His Ser Leu Lys Leu Leu Gln Ser Leu  
 130 135 140  
 Ala Pro Pro Gly Asp Lys Val Val Tyr Leu Cys Pro Ile Leu Val Phe  
 145 150 155 160  
 Val Ala Gln Arg Thr Leu Arg Val Ser Arg Val Thr Arg Leu Val Pro  
 165 170 175  
 Gln Lys Ile Ser Gly Asn Ile Thr Ala Ala Val Arg Met Leu Gln Ser  
 180 185 190  
 Leu Ser Thr Tyr Ala Val Pro Pro Glu Pro Gln Thr Arg Arg Ser Arg  
 195 200 205  
 25 Arg Arg Val Ala Ala Thr Ala Arg Pro Gln Arg Pro Pro Ser Pro Thr  
 210 215 220  
 Arg Asp Pro Glu Gly Thr Ala Gly His Pro Ala Pro Pro Glu Ser Asp  
 225 230 235 240  
 Pro Pro Ser Pro Gly Val Val Gly Val Ala Ala Glu Gly Gly Gly Val  
 245 250 255  
 30 Leu Gln Lys Ile Ala Ala Leu Phe Cys Val Pro Val Ala Ala Lys Ser  
 260 265 270  
 Arg Pro Arg Thr Lys Thr Glu  
 275  
 35

## (2) INFORMATION FOR SEQ ID NO:239:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 571 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

```

5  Met Asp Pro Tyr Tyr Pro Phe Asp Ala Leu Asp Val Trp Glu His Arg
   1             5             10             15
Arg Phe Ile Val Ala Asp Ser Arg Ser Phe Ile Thr Pro Glu Phe Pro
   20             25             30
Arg Asp Phe Trp Met Leu Pro Val Phe Asn Ile Pro Arg Glu Thr Ala
10      35             40             45
Ala Glu Arg Ala Ala Val Leu Gln Ala Gln Arg Thr Ala Ala Ala Ala
   50             55             60
Ala Leu Glu Asn Ala Ala Leu Gln Ala Ala Glu Leu Pro Val Asp Ile
65             70             75             80
15  Glu Arg Arg Ile Arg Pro Ile Glu Gln Gln Val His His Ile Ala Asp
   85             90             95
Ala Leu Glu Ala Leu Glu Thr Ala Ala Ala Ala Ala Glu Glu Ala Asp
   100            105            110
Ala Ala Arg Asp Ala Glu Arg Glu Gly Ala Ala Asp Gly Ala Ala Pro
20      115            120            125
Ser Pro Thr Ala Gly Pro Ala Ala Ala Glu Met Glu Val Gln Ile Val
   130            135            140
Arg Asn Asp Pro Pro Leu Arg Tyr Asp Thr Asn Leu Pro Val Asp Leu
145            150            155            160
25  Leu His Met Val Tyr Ala Gly Arg Gly Ala Ala Gly Ser Ser Gly Val
   165            170            175
Val Phe Gly Thr Trp Tyr Arg Thr Ile Gln Glu Arg Thr Ile Ala Asp
   180            185            190
Phe Pro Leu Thr Thr Arg Ser Ala Asp Phe Arg Asp Gly Arg Met Ser
30      195            200            205
Lys Thr Phe Met Thr Ala Leu Val Leu Ser Leu Gln Ser Cys Gly Arg
   210            215            220
Leu Tyr Val Gly Gln Arg His Tyr Ser Ala Phe Glu Cys Ala Val Leu
225            230            235            240
35  Cys Leu Tyr Leu Leu Tyr Arg Thr Thr His Glu Ser Ser Pro Asp Arg
   245            250            255
Asp Arg Ala Pro Val Ala Phe Gly Asp Leu Leu Ala Arg Leu Pro Arg
   260            265            270
Tyr Leu Ala Arg Leu Ala Ala Val Ile Gly Asp Glu Ser Gly Arg Pro
40      275            280            285
Gln Tyr Arg Tyr Arg Asp Asp Lys Leu Pro Lys Ala Gln Phe Ala Ala
   290            295            300
Ala Gly Gly Arg Tyr Glu His Gly Ala Thr His Val Val Ile Ala Thr

```

305                      310                      315                      320  
 Leu Val Arg His Gly Val Leu Pro Ala Ala Pro Gly Asp Val Pro Arg  
                          325                      330                      335  
 Asp Thr Ser Thr Arg Val Asn Pro Asp Asp Val Ala His Arg Asp Asp  
 5                      340                      345                      350  
 Val Asn Arg Ala Ala Ala Ala Phe Leu Arg His Asn Leu Phe Leu Trp  
                          355                      360                      365  
 Glu Asp Gln Thr Leu Leu Arg Ala Thr Ala Asn Thr Ile Thr Ala Val  
                          370                      375                      380  
 10 Leu Arg Arg Leu Leu Ala Asn Gly Asn Val Tyr Ala Asp Arg Leu Asp  
                          385                      390                      395                      400  
 Asn Arg Leu Gln Leu Gly Met Leu Ile Pro Gly Ala Val Pro Ala Glu  
                          405                      410                      415  
 Ala Ile Arg Ala Ser Gly Leu Asp Ser Gly Ala Ile Lys Ser Gly Asp  
 15                      420                      425                      430  
 Asn Asn Leu Glu Ala Leu Cys Val Asn Tyr Val Leu Pro Leu Tyr Gln  
                          435                      440                      445  
 Ala Asp Pro Thr Val Glu Leu Thr Gln Leu Phe Pro Gly Leu Ala Ala  
                          450                      455                      460  
 20 Leu Cys Leu Asp Ala Gln Ala Gly Arg Pro Leu Ala Ser Thr Arg Arg  
                          465                      470                      475                      480  
 Val Val Asp Met Ser Ser Gly Ala Arg Gln Ala Ala Leu Val Arg Leu  
                          485                      490                      495  
 Thr Ala Leu Glu Leu Ile Asn Arg Thr Arg Thr Asn Thr Thr Pro Val  
 25                      500                      505                      510  
 Gly Glu Ile Ile Asn Ala His Asp Ala Leu Gly Ile Gln Tyr Glu Gln  
                          515                      520                      525  
 Gly Leu Gly Leu Leu Ala Gln Gln Ala Arg Ile Gln Ala Lys Arg Phe  
                          530                      535                      540  
 30 Ala Thr Phe Asn Val Gly Ser Asp Tyr Asp Leu Leu Tyr Phe Leu Cys  
                          545                      550                      555                      560  
 Leu Gly Phe Ile Pro Gln Tyr Leu Ser Val Ala  
                          565                      570

35                      (2) INFORMATION FOR SEQ ID NO:240:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 651 amino acids  
 (B) TYPE: amino acid  
 40 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

	Met	Ala	Ser	Ala	Glu	Met	Arg	Glu	Arg	Leu	Glu	Ala	Pro	Leu	Pro	Asp
5	1				5					10					15	
	Arg	Ala	Val	Pro	Ile	Tyr	Val	Ala	Gly	Phe	Leu	Ala	Leu	Tyr	Asp	Ser
				20					25					30		
	Gly	Asp	Pro	Gly	Glu	Leu	Ala	Leu	Asp	Pro	Asp	Thr	Val	Arg	Ala	Ala
			35					40					45			
10	Leu	Pro	Pro	Glu	Asn	Pro	Leu	Pro	Ile	Asn	Val	Asp	His	Arg	Ala	Arg
			50				55					60				
	Cys	Glu	Val	Gly	Arg	Val	Leu	Ala	Val	Val	Asn	Asp	Pro	Arg	Gly	Pro
	65				70						75				80	
	Phe	Phe	Val	Gly	Leu	Ile	Ala	Cys	Val	Gln	Leu	Glu	Arg	Val	Leu	Glu
15					85					90					95	
	Thr	Ala	Ala	Ser	Ala	Ala	Ile	Phe	Glu	Arg	Arg	Gly	Pro	Ala	Leu	Ser
					100					105				110		
	Arg	Glu	Glu	Arg	Leu	Leu	Tyr	Leu	Ile	Thr	Asn	Tyr	Leu	Pro	Ser	Val
				115					120				125			
20	Ser	Leu	Ser	Thr	Lys	Arg	Arg	Gly	Asp	Glu	Val	Pro	Pro	Asp	Arg	Thr
				130				135					140			
	Leu	Phe	Ala	His	Val	Cys	Ala	Ile	Gly	Arg	Arg	Leu	Gly	Thr	Ile	Val
	145				150							155			160	
	Thr	Tyr	Asp	Thr	Ser	Leu	Asp	Ala	Ala	Ile	Ala	Pro	Phe	Arg	His	Leu
25					165					170				175		
	Asp	Pro	Ala	Thr	Arg	Glu	Gly	Val	Arg	Arg	Glu	Ala	Ala	Glu	Ala	Glu
				180					185					190		
	Leu	Ala	Gly	Arg	Thr	Trp	Ala	Pro	Gly	Val	Glu	Ala	Leu	Thr	His	Thr
			195				200						205			
30	Leu	Leu	Ser	Thr	Ala	Val	Asn	Asn	Met	Met	Leu	Arg	Asp	Arg	Trp	Ser
				210			215					220				
	Leu	Val	Ala	Glu	Arg	Arg	Arg	Gln	Ala	Gly	Ile	Ala	Gly	His	Thr	Tyr
	225				230						235				240	
	Leu	Gln	Ala	Ser	Glu	Lys	Phe	Lys	Ile	Trp	Gly	Ala	Glu	Ser	Ala	Pro
35					245					250				255		
	Ala	Pro	Glu	Arg	Gly	Tyr	Lys	Thr	Gly	Ala	Pro	Gly	Ala	Met	Asp	Thr
				260					265				270			
	Ser	Pro	Ala	Ala	Ser	Val	Pro	Ala	Pro	Gln	Val	Ala	Val	Arg	Ala	Arg
			275					280					285			
40	Gln	Val	Ala	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Phe	Pro	Ala	Pro	Ala
			290				295					300				
	Asp	Met	Asn	Pro	Val	Ser	Ala	Ser	Gly	Ala	Pro	Ala	Pro	Pro	Pro	Pro
	305					310					315				320	

Gly Asp Gly Ser Tyr Leu Trp Ile Pro Ala Phe His Tyr Asn Gln Leu  
 325 330 335  
 Val Thr Gly Gln Ser Ala Pro His His Pro Pro Leu Thr Ala Cys Gly  
 340 345 350  
 5 Leu Pro Ala Ala Gly Thr Val Ala Tyr Gly His Pro Gly Ala Gly Pro  
 355 360 365  
 Ser Pro His Tyr Pro Pro Pro Pro Ala His Pro Tyr Pro Gly Met Leu  
 370 375 380  
 Phe Ala Gly Pro Ser Pro Leu Glu Ala Gln Ile Ala Ala Leu Val Gly  
 10 385 390 395 400  
 Ala Ile Ala Ala Asp Arg Gln Ala Gly Gly Leu Pro Ala Ala Ala Gly  
 405 410 415  
 Asp His Gly Ile Arg Gly Ser Ala Lys Arg Arg Arg His Glu Val Glu  
 420 425 430  
 15 Gln Pro Glu Tyr Asp Cys Gly Arg Asp Glu Pro Asp Arg Asp Phe Pro  
 435 440 445  
 Tyr Tyr Pro Gly Glu Ala Arg Pro Glu Pro Arg Pro Val Asp Ser Arg  
 450 455 460  
 Arg Ala Ala Arg Gln Ala Ser Gly Phe Thr Ile Thr Ala Leu Val Gly  
 20 465 470 475 480  
 Ala Val Thr Ser Leu Gln Gln Glu Leu Ala His Met Arg Ala Arg Thr  
 485 490 495  
 His Ala Pro Tyr Gly Pro Tyr Pro Pro Val Gly Pro Tyr His His Pro  
 500 505 510  
 25 His Ala Asp Thr Glu Thr Pro Ala Gln Pro Pro Arg Tyr Pro Ala Glu  
 515 520 525  
 Ala Val Tyr Leu Pro Pro Pro His Ile Ala Pro Pro Gly Pro Pro Leu  
 530 535 540  
 Ser Gly Ala Val Pro Pro Pro Ser Tyr Pro Pro Val Ala Val Thr Pro  
 30 545 550 555 560  
 Gly Pro Ala Pro Pro Leu His Gln Pro Ser Pro Ala His Ala His Pro  
 565 570 575  
 Pro Pro Pro Pro Pro Gly Pro Thr Pro Pro Pro Ala Ala Ser Leu Pro  
 580 585 590  
 35 Gln Pro Glu Ala Pro Gly Ala Glu Ala Gly Ala Leu Val Asn Ala Ser  
 595 600 605  
 Ser Ala Ala His Val Lys Arg Gly His Gly Pro Gly Arg Arg Ser Val  
 610 615 620  
 Cys Val Thr Asp Asp Gly Val Pro Leu Thr Arg Leu Gln Asp Pro Asp  
 40 625 630 635 640  
 Leu Gly Gly Val Cys Val Phe Ile Tyr Phe Lys  
 645 650

## (2) INFORMATION FOR SEQ ID NO:241:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 896 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

Met Arg Gly Gly Gly Leu Ile Cys Ala Leu Val Val Gly Ala Leu Val  
 1 5 10 15  
 15 Ala Ala Val Ala Ser Ala Ala Pro Ala Ala Pro Ala Ala Pro Arg Ala  
 20 25 30  
 Ser Gly Gly Val Ala Ala Thr Val Ala Ala Asn Gly Gly Pro Ala Ser  
 35 40 45  
 Arg Pro Pro Pro Val Pro Ser Pro Ala Thr Thr Lys Ala Arg Lys Arg  
 20 50 55 60  
 Lys Thr Lys Lys Pro Pro Lys Arg Pro Glu Ala Thr Pro Pro Pro Asp  
 65 70 75 80  
 Ala Asn Ala Thr Val Ala Ala Gly His Ala Thr Leu Arg Ala His Leu  
 85 90 95  
 25 Arg Glu Ile Lys Val Glu Asn Ala Asp Ala Gln Phe Tyr Val Cys Pro  
 100 105 110  
 Pro Pro Thr Gly Ala Thr Val Val Gln Phe Glu Gln Pro Arg Arg Cys  
 115 120 125  
 Pro Trp Glu Gly Gln Asn Tyr Thr Glu Gly Ile Ala Val Val Phe Lys  
 30 130 135 140  
 Glu Asn Ile Ala Pro Tyr Lys Phe Lys Ala Thr Met Tyr Tyr Lys Asp  
 145 150 155 160  
 Val Thr Val Ser Gln Val Trp Phe Gly His Arg Tyr Ser Gln Phe Met  
 165 170 175  
 35 Gly Ile Phe Glu Asp Arg Ala Pro Val Pro Phe Glu Glu Val Ile Asp  
 180 185 190  
 Lys Ile Asn Ala Lys Gly Val Cys Arg Ser Thr Ala Lys Tyr Val Arg  
 195 200 205  
 Asn Asn Met Thr Ala Phe His Arg Asp Asp His Glu Thr Asp Met Glu  
 40 210 215 220  
 Leu Lys Pro Ala Lys Val Ala Thr Arg Thr Ser Arg Gly Trp His Thr  
 225 230 235 240  
 Thr Asp Leu Lys Tyr Asn Pro Ser Arg Val Glu Ala Phe His Arg Tyr

		245		250		255										
	Gly	Thr	Thr	Val	Asn	Cys	Ile	Val	Glu	Glu	Val	Asp	Ala	Arg	Ser	Val
				260					265							270
	Tyr	Pro	Tyr	Asp	Glu	Phe	Val	Leu	Ala	Thr	Gly	Asp	Phe	Val	Tyr	Met
5			275					280						285		
	Ser	Pro	Phe	Tyr	Gly	Tyr	Arg	Glu	Gly	Ser	His	Thr	Glu	His	Thr	Ser
		290					295						300			
	Tyr	Ala	Ala	Asp	Arg	Phe	Lys	Gln	Val	Asp	Gly	Phe	Tyr	Ala	Arg	Asp
	305					310					315				320	
10	Leu	Thr	Thr	Lys	Ala	Arg	Ala	Thr	Ser	Pro	Thr	Thr	Arg	Asn	Leu	Leu
				325						330					335	
	Thr	Thr	Pro	Lys	Phe	Thr	Val	Ala	Trp	Asp	Trp	Val	Pro	Lys	Arg	Pro
				340					345					350		
	Ala	Val	Cys	Thr	Met	Thr	Lys	Trp	Gln	Glu	Val	Asp	Glu	Met	Leu	Arg
15			355					360					365			
	Ala	Glu	Tyr	Gly	Gly	Ser	Phe	Arg	Phe	Ser	Ser	Asp	Ala	Ile	Ser	Thr
	370					375						380				
	Thr	Phe	Thr	Thr	Asn	Leu	Thr	Gln	Tyr	Ser	Leu	Ser	Arg	Val	Asp	Leu
	385				390						395				400	
20	Gly	Asp	Cys	Ile	Gly	Arg	Asp	Ala	Arg	Glu	Ala	Ile	Asp	Arg	Met	Phe
				405						410					415	
	Ala	Arg	Lys	Tyr	Asn	Ala	Thr	His	Ile	Lys	Val	Gly	Gln	Pro	Gln	Tyr
			420					425					430			
	Tyr	Leu	Ala	Thr	Gly	Gly	Phe	Leu	Ile	Ala	Tyr	Gln	Pro	Leu	Leu	Ser
25			435					440					445			
	Asn	Thr	Leu	Ala	Glu	Leu	Tyr	Val	Arg	Glu	Tyr	Met	Arg	Glu	Gln	Asp
	450					455						460				
	Arg	Lys	Pro	Arg	Asn	Ala	Thr	Pro	Ala	Pro	Leu	Arg	Glu	Ala	Pro	Ser
	465				470					475				480		
30	Ala	Asn	Ala	Ser	Val	Glu	Arg	Ile	Lys	Thr	Thr	Ser	Ser	Ile	Glu	Phe
				485						490				495		
	Ala	Arg	Leu	Gln	Phe	Thr	Tyr	Asn	His	Ile	Gln	Arg	His	Val	Asn	Asp
			500					505					510			
	Met	Leu	Gly	Arg	Ile	Ala	Val	Ala	Trp	Cys	Glu	Leu	Gln	Asn	His	Glu
35			515					520					525			
	Leu	Thr	Leu	Trp	Asn	Glu	Ala	Arg	Lys	Leu	Asn	Pro	Asn	Ala	Ile	Ala
	530					535						540				
	Ser	Ala	Thr	Val	Gly	Arg	Arg	Val	Ser	Ala	Arg	Met	Leu	Gly	Asp	Val
	545				550					555				560		
40	Met	Ala	Val	Ser	Thr	Cys	Val	Pro	Val	Ala	Pro	Asp	Asn	Val	Ile	Val
				565						570				575		
	Gln	Asn	Ser	Met	Arg	Val	Ser	Ser	Arg	Pro	Gly	Thr	Cys	Arg	Pro	Leu
				580					585					590		
									609							



Val Ser Phe Arg Tyr Glu Asp Gln Gly Pro Leu Ile Glu Gly Gln Leu  
 595 600 605  
 Gly Glu Asn Asn Glu Leu Arg Leu Thr Arg Asp Ala Leu Glu Pro Cys  
 610 615 620  
 5 Thr Val Gly His Arg Arg Tyr Phe Ile Phe Gly Gly Gly Tyr Val Tyr  
 625 630 635 640  
 Phe Glu Glu Tyr Ala Tyr Ser His Gln Leu Ser Arg Ala Asp Val Thr  
 645 650 655  
 10 Thr Val Ser Thr Phe Ile Asp Leu Asn Ile Thr Met Leu Glu Asp His  
 660 665 670  
 Glu Phe Val Pro Leu Glu Val Tyr Thr Arg His Glu Ile Lys Asp Ser  
 675 680 685  
 Gly Leu Leu Asp Tyr Thr Glu Val Gln Arg Arg Asn Gln Leu His Asp  
 690 695 700  
 15 Leu Arg Phe Ala Asp Ile Asp Thr Val Ile Arg Ala Asp Ala Asn Ala  
 705 710 715 720  
 Ala Met Phe Ala Gly Leu Cys Ala Phe Phe Glu Gly Met Gly Asp Leu  
 725 730 735  
 Gly Arg Ala Val Gly Lys Val Val Met Gly Val Val Gly Gly Val Val  
 20 740 745 750  
 Ser Ala Val Ser Gly Val Ser Ser Phe Met Ser Asn Pro Phe Gly Ala  
 755 760 765  
 Val Gly Leu Leu Val Leu Ala Gly Leu Val Ala Ala Phe Phe Ala Phe  
 770 775 780  
 25 Arg Tyr Val Leu Gln Leu Gln Arg Asn Pro Met Lys Ala Leu Tyr Pro  
 785 790 795 800  
 Leu Thr Thr Lys Glu Leu Lys Thr Ser Asp Pro Gly Gly Val Gly Gly  
 805 810 815  
 Glu Gly Glu Glu Gly Ala Glu Gly Gly Gly Phe Asp Glu Ala Lys Leu  
 30 820 825 830  
 Ala Glu Ala Arg Glu Met Ile Arg Tyr Met Ala Leu Val Ser Ala Met  
 835 840 845  
 Glu Arg Thr Glu His Lys Ala Arg Lys Lys Gly Thr Ser Ala Leu Leu  
 850 855 860  
 35 Ser Ser Lys Val Thr Asn Met Val Leu Arg Lys Arg Asn Lys Ala Arg  
 865 870 875 880  
 Tyr Ser Pro Leu His Asn Glu Asp Glu Ala Gly Asp Glu Asp Glu Leu  
 885 890 895

40 (2) INFORMATION FOR SEQ ID NO:242:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 69 amino acids

610

(B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

```

Val Val Ala Gly Leu Gly Thr Gly Gly Gly Arg Glu Ala Gly Pro Pro
10 1 5 10 15
Phe Ala Ala Thr Val Ala Ala Thr Pro Pro Glu Arg Ala Ala Gly Ala
20 25 30
Ala Gly Ala Ala Asp Ala Thr Ala Ala Thr Ser Ala Pro Thr Thr Ser
35 40 45
15 Ala Gln Ile Lys Pro Pro Pro Arg Met Ala Gly Leu Arg Gly Arg Val
50 55 60
Ala Pro Ala Ala Arg
65

```

20 (2) INFORMATION FOR SEQ ID NO:243:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 773 amino acids  
 (B) TYPE: amino acid  
 25 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:

```

Met Ala Ala Ala Pro Pro Ala Ala Val Ser Glu Pro Thr Ala Ala Arg
1 5 10 15
Gln Lys Leu Leu Ala Leu Leu Gly Gln Val Gln Thr Tyr Val Phe Gln
35 20 25 30
Leu Glu Leu Leu Arg Arg Cys Asp Pro Gln Ile Gly Leu Gly Lys Leu
35 40 45
Ala Gln Leu Lys Leu Asn Ala Leu Gln Val Arg Val Leu Arg Arg His
50 55 60
40 Leu Arg Pro Gly Leu Glu Ala Gln Ala Ala Ala Phe Leu Thr Pro Leu
65 70 75 80
Ser Val Thr Leu Glu Leu Leu Glu Tyr Ala Trp Arg Glu Gly Glu
85 90 95

```

	Arg	Leu	Leu	Gly	His	Leu	Glu	Thr	Phe	Ala	Thr	Thr	Gly	Asp	Val	Ser
				100					105					110		
	Ala	Phe	Phe	Thr	Glu	Thr	Met	Gly	Leu	Ala	Arg	Pro	Cys	Pro	Tyr	His
			115					120					125			
5	Gln	Gln	Ile	Arg	Leu	Glu	Thr	Tyr	Gly	Gly	Asp	Val	Arg	Met	Glu	Leu
		130					135					140				
	Cys	Phe	Leu	His	Asp	Val	Glu	Asn	Phe	Leu	Lys	Gln	Leu	Asn	Tyr	Cys
	145				150					155					160	
	His	Leu	Ile	Thr	Pro	Pro	Ser	Gly	Ala	Thr	Ala	Ala	Leu	Glu	Arg	Val
10				165					170					175		
	Arg	Glu	Phe	Met	Val	Ala	Ala	Val	Gly	Ser	Gly	Leu	Ile	Val	Pro	Pro
				180					185				190			
	Glu	Leu	Ser	Asp	Pro	Ser	His	Pro	Cys	Ala	Val	Cys	Phe	Glu	Glu	Leu
		195					200				205					
15	Cys	Val	Thr	Ala	Asn	Gln	Gly	Ala	Thr	Ile	Ala	Arg	Arg	Leu	Ala	Asp
		210				215					220					
	Arg	Ile	Cys	Asn	His	Val	Thr	Gln	Gln	Ala	Gln	Val	Arg	Leu	Asp	Ala
	225				230					235					240	
	Asn	Glu	Leu	Arg	Arg	Tyr	Leu	Pro	His	Ala	Ala	Gly	Leu	Ser	Asp	Ala
20				245						250				255		
	Ala	Arg	Ala	Arg	Ala	Leu	Cys	Val	Leu	Asp	Gln	Ala	Arg	Thr	Ala	Ala
				260					265				270			
	Gly	Gly	Gly	Ala	Arg	Ala	Gly	Pro	Pro	Pro	Ala	Asp	Ser	Ser	Ser	Val
		275					280				285					
25	Arg	Glu	Glu	Ala	Asp	Ala	Leu	Leu	Glu	Ala	His	Asp	Val	Phe	Gln	Ala
		290				295					300					
	Thr	Thr	Pro	Gly	Ala	Ile	Ser	Glu	Leu	Arg	Phe	Trp	Leu	Ala	Ser	Gly
	305				310					315					320	
	Asp	Arg	Ala	Arg	His	Ser	Thr	Met	Asp	Ala	Phe	Ala	Asp	Asn	Leu	Asn
30				325						330				335		
	Ala	Gln	Arg	Glu	Leu	Gln	Gln	Glu	Thr	Ala	Ala	Val	Ala	Val	Glu	Leu
		340						345				350				
	Ala	Leu	Phe	Gly	Arg	Arg	Ala	Glu	His	Phe	Asp	Arg	Ala	Phe	Gly	Gly
		355				360					365					
35	His	Leu	Ala	Ala	Leu	Asp	Met	Val	Asp	Ala	Leu	Ile	Ile	Gly	Gly	Gln
		370				375					380					
	Ala	Thr	Ser	Pro	Asp	Asp	Gln	Ile	Glu	Ala	Leu	Ile	Arg	Ala	Cys	Tyr
	385				390					395					400	
	Asp	His	His	Leu	Thr	Thr	Pro	Leu	Leu	Arg	Arg	Leu	Val	Ser	Pro	Glu
40				405						410				415		
	Gln	Cys	Asp	Glu	Glu	Ala	Leu	Arg	Arg	Val	Leu	Ala	Arg	Leu	Gly	Ala
			420					425				430				
	Gly	Gly	Ala	Thr	Gly	Gly	Ala	Glu	Glu	Glu	Pro	Arg	Ala	Ala	Ala	

435                      440                      445  
 Glu Glu Gly Gly Arg Arg Arg Gly Ala Gly Thr Pro Ala Ser Glu Asp  
 450                      455                      460  
 Gly Glu Arg Gly Pro Glu Pro Gly Ala Gln Gly Pro Glu Ser Trp Gly  
 5    465                      470                      475                      480  
 Asp Ile Ala Thr Arg Ala Ala Ala Asp Val Pro Glu Arg Arg Arg Leu  
 485                      490                      495  
 Tyr Ala Asp Arg Leu Thr Lys Arg Ser Leu Ala Ser Leu Gly Arg Cys  
 500                      505                      510  
 10 Val Arg Glu Gln Arg Gly Glu Leu Glu Lys Met Leu Arg Val Ser Val  
 515                      520                      525  
 His Gly Glu Val Leu Pro Ala Thr Phe Ala Ala Val Ala Asn Gly Phe  
 530                      535                      540  
 Ala Ala Arg Ala Arg Phe Cys Ala Leu Thr Ala Gly Ala Gly Thr Val  
 15 545                      550                      555                      560  
 Ile Asp Asn Arg Ala Ala Pro Gly Val Phe Asp Ala His Arg Phe Met  
 565                      570                      575  
 Arg Ala Ser Leu Leu Arg His Gln Val Asp Pro Ala Leu Leu Pro Ser  
 580                      585                      590  
 20 Ile Thr Phe Phe Glu Leu Val Asn Gly Pro Leu Phe Asp His Ser Thr  
 595                      600                      605  
 His Ser Phe Ala Gln Pro Pro Asn Thr Ala Leu Tyr Tyr Ser Val Glu  
 610                      615                      620  
 Asn Val Gly Leu Leu Pro His Leu Lys Glu Glu Leu Ala Arg Phe Ile  
 25 625                      630                      635                      640  
 Met Gly Ala Gly Gly Ser Gly Ala Asp Trp Ala Val Ser Glu Phe Gln  
 645                      650                      655  
 Lys Phe Tyr Cys Phe Asp Gly Val Ser Gly Ile Thr Pro Thr Gln Arg  
 660                      665                      670  
 30 Ala Ala Trp Arg Tyr Ile Arg Glu Leu Ile Ile Ala Thr Thr Leu Phe  
 675                      680                      685  
 Ala Ser Val Tyr Arg Cys Gly Glu Leu Glu Leu Arg Arg Pro Asp Cys  
 690                      695                      700  
 Ser Arg Pro Thr Ser Glu Gly Arg Tyr Pro Pro Gly Val Tyr Leu Thr  
 35 705                      710                      715                      720  
 Tyr Asn Ser Asp Cys Pro Leu Val Ala Ile Val Glu Ser Gly Pro Asp  
 725                      730                      735  
 Gly Cys Ile Gly Pro Arg Ser Val Val Val Tyr Asp Arg Asp Val Phe  
 740                      745                      750  
 40 Ser Ile Lys Val Leu Gln His Leu Ala Pro Arg Leu Ala Gly Gly Gly  
 755                      760                      765  
 Ser Asp Ala Pro Pro  
 770

## (2) INFORMATION FOR SEQ ID NO:244:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 616 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 10 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:

Met Asp Thr Lys Pro Lys Thr Thr Thr Thr Val Lys Val Pro Pro Gly  
 15 1 5 10 15  
 Pro Met Gly Tyr Val Tyr Gly Arg Ala Cys Pro Ala Glu Gly Leu Glu  
 20 20 25 30  
 Leu Leu Ser Leu Leu Ser Ala Arg Ser Gly Asp Ala Asp Val Ala Val  
 35 40 45  
 20 Ala Pro Leu Ile Val Gly Leu Thr Val Glu Ser Gly Phe Glu Ala Asn  
 50 55 60  
 Val Ala Ala Val Val Gly Ser Arg Thr Thr Gly Leu Gly Gly Thr Ala  
 65 70 75 80  
 Val Ser Leu Lys Leu Met Pro Ser His Tyr Ser Pro Ser Val Tyr Val  
 25 85 90 95  
 Phe His Gly Gly Arg His Leu Ala Pro Ser Thr Gln Ala Pro Asn Leu  
 100 105 110  
 Thr Arg Leu Cys Glu Arg Ala Arg Arg His Phe Gly Phe Ser Asp Tyr  
 115 120 125  
 30 Ala Pro Arg Pro Cys Asp Leu Lys His Glu Thr Thr Gly Asp Ala Leu  
 130 135 140  
 Cys Glu Arg Leu Gly Leu Asp Pro Asp Arg Ala Leu Leu Tyr Leu Val  
 145 150 155 160  
 Ile Thr Glu Gly Phe Arg Glu Ala Val Cys Ile Ser Asn Thr Phe Leu  
 35 165 170 175  
 His Leu Gly Gly Met Asp Lys Val Thr Ile Gly Asp Ala Glu Val His  
 180 185 190  
 Arg Ile Pro Val Tyr Pro Leu Gln Met Phe Met Pro Asp Phe Ser Arg  
 195 200 205  
 40 Val Ile Ala Asp Pro Phe Asn Cys Asn His Arg Ser Ile Gly Glu Asn  
 210 215 220  
 Phe Asn Tyr Pro Leu Pro Phe Phe Asn Arg Pro Leu Ala Arg Leu Leu  
 225 230 235 240

[illegible]

			580					585	.			590					
Gly	Arg	Leu	Glu	Thr	Ile	Ile	Gly	Thr	Arg	Glu	Ala	Leu	His	Thr	Val		
			595					600				605					
Val	Asn	Asn	Ile	Lys	Gln	Leu	Val										
5			610					615									

(2) INFORMATION FOR SEQ ID NO:245:

(i) SEQUENCE CHARACTERISTICS:

10 (A) LENGTH: 616 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

[illegible]

	195	200	205
	Val Ile Ala Asp Pro Phe Asn Cys Asn His Arg Ser Ile Gly Glu Asn		
	210	215	220
5	Phe Asn Tyr Pro Leu Pro Phe Phe Asn Arg Pro Leu Ala Arg Leu Leu		
	225	230	235
	Phe Glu Ala Val Val Gly Pro Ala Ala Val Arg Ala Arg Asn Val Asp		
	245	250	255
	Ala Val Ala Arg Ala Ala Ala His Leu Ala Phe Asp Glu Asn His Glu		
	260	265	270
10	Gly Ala Ala Leu Pro Ala Asp Ile Thr Phe Thr Ala Phe Glu Ala Ser		
	275	280	285
	Gln Gly Lys Pro Gln Arg Gly Ala Arg Asp Ala Gly Asn Lys Gly Pro		
	290	295	300
	Ala Gly Gly Phe Glu Gln Arg Leu Ala Ser Val Met Ala Gly Asp Ala		
15	305	310	315
	Ala Leu Glu Ser Ile Val Ser Met Ala Val Phe Asp Glu Pro Pro Pro		
	325	330	335
	Asp Ile Thr Thr Trp Pro Leu Leu Glu Gly Gln Glu Thr Pro Ala Ala		
	340	345	350
20	Arg Ala Gly Ala Val Gly Ala Tyr Leu Ala Arg Ala Ala Gly Leu Val		
	355	360	365
	Gly Ala Met Val Phe Ser Thr Asn Ser Ala Leu His Leu Thr Glu Val		
	370	375	380
	Asp Asp Ala Gly Pro Ala Asp Pro Lys Asp His Ser Lys Pro Ser Phe		
25	385	390	395
	Tyr Arg Phe Phe Leu Val Pro Gly Thr His Val Ala Ala Asn Pro Gln		
	405	410	415
	Leu Asp Arg Glu Gly His Val Val Pro Gly Tyr Glu Gly Arg Pro Thr		
	420	425	430
30	Ala Pro Leu Val Gly Gly Thr Gln Glu Phe Ala Gly Glu His Leu Ala		
	435	440	445
	Met Leu Cys Gly Phe Ser Pro Ala Leu Leu Ala Lys Met Leu Phe Tyr		
	450	455	460
	Leu Glu Arg Cys Asp Gly Gly Val Ile Val Gly Arg Gln Glu Met Asp		
35	465	470	475
	Val Phe Arg Tyr Val Ala Asp Ser Gly Gln Thr Asp Val Pro Cys Asn		
	485	490	495
	Leu Cys Thr Phe Glu Thr Arg His Ala Cys Ala His Thr Thr Leu Met		
	500	505	510
40	Arg Leu Arg Ala Arg His Pro Lys Phe Ala Ser Ala Arg Ala Ile Gly		
	515	520	525
	Val Phe Gly Thr Met Asn Ser Ala Tyr Ser Asp Cys Asp Val Leu Gly		
	530	535	540



	Asn	Tyr	Ala	Ala	Phe	Ser	Ala	Leu	Lys	Arg	Ala	Asp	Gly	Ser	Glu	Asn
	545					550					555					560
	Thr	Arg	Thr	Ile	Met	Gln	Glu	Tyr	Ala	Ala	Thr	Glu	Arg	Val	Met	Ala
					565					570						575
5	Glu	Leu	Glu	Ala	Leu	Gln	Tyr	Val	Asp	Gln	Ala	Val	Pro	Thr	Ala	Leu
					580					585					590	
	Gly	Arg	Leu	Glu	Thr	Ile	Ile	Gly	Thr	Arg	Glu	Ala	Leu	His	Thr	Val
					595				600				605			
	Val	Asn	Asn	Ile	Lys	Gln	Leu	Val								
10																
		610							615							

(2) INFORMATION FOR SEQ ID NO:246:

(i) SEQUENCE CHARACTERISTICS:

15           (A) LENGTH: 1228 amino acids  
             (B) TYPE: amino acid  
             (C) STRANDEDNESS: single  
             (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEO ID NO:246:

25	Met	Phe	Cys	Ala	Ala	Gly	Gly	Pro	Thr	Ser	Pro	Gly	Gly	Lys	Ser	Ala
	1				5					10					15	
	Ala	Arg	Ala	Ala	Ser	Gly	Phe	Phe	Ala	Pro	His	Asn	Pro	Arg	Gly	Ala
				20					25					30		
	Thr	Gln	Thr	Ala	Pro	Pro	Pro	Cys	Arg	Arg	Gln	Asn	Phe	Tyr	Asn	Pro
			35					40					45			
30	His	Leu	Ala	Gln	Thr	Gly	Thr	Gln	Pro	Lys	Ala	Pro	Gly	Pro	Ala	Gln
		50					55					60				
	Arg	His	Thr	Tyr	Tyr	Ser	Glu	Cys	Asp	Glu	Phe	Arg	Phe	Ile	Ala	Pro
	65					70				75					80	
	Arg	Ser	Leu	Asp	Glu	Asp	Ala	Pro	Ala	Glu	Gln	Arg	Thr	Gly	Val	His
35				85						90					95	
	Asp	Gly	Arg	Leu	Arg	Arg	Ala	Pro	Lys	Val	Tyr	Cys	Gly	Gly	Asp	Glu
				100					105					110		
	Arg	Asp	Val	Leu	Arg	Val	Gly	Pro	Glu	Gly	Phe	Trp	Pro	Arg	Arg	Leu
			115					120					125			
40	Arg	Leu	Trp	Gly	Gly	Ala	Asp	His	Ala	Pro	Glu	Gly	Phe	Asp	Pro	Thr
		130					135					140				
	Val	Thr	Val	Phe	His	Val	Tyr	Asp	Ile	His	Val	Glu	His	Ala	Tyr	Ser
	145					150					155					160

Met Arg Ala Ala Gln Leu His Glu Arg Phe Met Asp Ala Ile Thr Pro  
 165 170 175  
 Ala Gly Thr Val Ile Thr Leu Leu Gly Leu Thr Pro Glu Gly His Arg  
 180 185 190  
 5 Val Ala Val His Val Tyr Gly Thr Arg Gln Tyr Phe Tyr Met Asn Lys  
 195 200 205  
 Ala Glu Val Asp Arg His Leu Gln Cys Arg Ala Pro Arg Asp Leu Cys  
 210 215 220  
 10 Glu Arg Leu Ala Ala Ala Leu Arg Glu Ser Pro Gly Ala Ser Phe Arg  
 225 230 235 240  
 Gly Ile Ser Ala Asp His Phe Glu Ala Glu Val Val Glu Arg Ala Asp  
 245 250 255  
 Val Tyr Tyr Tyr Glu Trp Thr Leu Tyr Tyr Arg Val Phe Val Arg Ser  
 260 265 270  
 15 Gly Arg Ala Tyr Leu Cys Asp Asn Phe Cys Pro Ala Ile Arg Lys Tyr  
 275 280 285  
 Glu Gly Gly Val Asp Ala Thr Thr Arg Phe Ile Leu Asp Asn Pro Gly  
 290 295 300  
 Phe Val Thr Phe Gly Trp Tyr Arg Leu Lys Pro Gly Arg Gly Asn Ala  
 20 305 310 315 320  
 Pro Ala Gln Pro Arg Pro Pro Thr Ala Phe Gly Thr Ser Ser Asp Val  
 325 330 335  
 Glu Phe Asn Cys Thr Ala Asp Asn Leu Ala Val Glu Gly Ala Met Cys  
 340 345 350  
 25 Asp Leu Pro Ala Tyr Lys Leu Met Cys Phe Asp Ile Glu Cys Lys Ala  
 355 360 365  
 Gly Gly Glu Asp Glu Leu Ala Phe Pro Val Ala Glu Arg Pro Glu Asp  
 370 375 380  
 Leu Val Ile Gln Ile Ser Cys Leu Leu Tyr Asp Leu Ser Thr Thr Ala  
 30 385 390 395 400  
 Leu Glu His Ile Leu Leu Phe Ser Leu Gly Ser Cys Asp Leu Pro Glu  
 405 410 415  
 Ser His Leu Ser Asp Leu Ala Ser Arg Gly Leu Pro Ala Pro Val Val  
 420 425 430  
 35 Leu Glu Phe Asp Ser Glu Phe Glu Met Leu Leu Ala Phe Met Thr Phe  
 435 440 445  
 Val Lys Gln Tyr Gly Pro Glu Phe Val Thr Gly Tyr Asn Ile Ile Asn  
 450 455 460  
 Phe Asp Trp Pro Phe Val Leu Thr Lys Leu Thr Glu Ile Tyr Lys Val  
 40 465 470 475 480  
 Pro Leu Asp Gly Tyr Gly Arg Met Asn Gly Arg Gly Val Phe Arg Val  
 485 490 495  
 Trp Asp Ile Gly Gln Ser His Phe Gln Lys Arg Ser Lys Ile Lys Val

	500		505		510
	Asn Gly Met Val Asn Ile Asp Met Tyr Gly Ile Ile Thr Asp Lys Val				
	515		520		525
	Lys Leu Ser Ser Tyr Lys Leu Asn Ala Val Ala Glu Ala Val Leu Lys				
5	530		535		540
	Asp Lys Lys Lys Asp Leu Ser Tyr Arg Asp Ile Pro Ala Tyr Tyr Ala				
	545		550		555 560
	Ser Gly Pro Ala Gln Arg Gly Val Ile Gly Glu Tyr Cys Val Gln Asp				
	565		570		575
10	Ser Leu Leu Val Gly Gln Leu Phe Phe Lys Phe Leu Pro His Leu Glu				
	580		585		590
	Leu Ser Ala Val Ala Arg Leu Ala Gly Ile Asn Ile Thr Arg Thr Ile				
	595		600		605
	Tyr Asp Gly Gln Gln Ile Arg Val Phe Thr Cys Leu Leu Arg Leu Ala				
15	610		615		620
	Gly Gln Lys Gly Phe Ile Leu Pro Asp Thr Gln Gly Arg Phe Arg Gly				
	625		630		635 640
	Leu Asp Lys Glu Ala Pro Lys Arg Pro Ala Val Pro Arg Gly Glu Gly				
	645		650		655
20	Glu Arg Pro Gly Asp Gly Asn Gly Asp Glu Asp Lys Asp Asp Asp Glu				
	660		665		670
	Asp Gly Asp Glu Asp Gly Asp Glu Arg Glu Glu Val Ala Arg Glu Thr				
	675		680		685
	Gly Gly Arg His Val Gly Tyr Gln Gly Ala Arg Val Leu Asp Pro Thr				
25	690		695		700
	Ser Gly Phe His Val Asp Pro Val Val Val Phe Asp Phe Ala Ser Leu				
	705		710		715 720
	Tyr Pro Ser Ile Ile Gln Ala His Asn Leu Cys Phe Ser Thr Leu Ser				
	725		730		735
30	Leu Arg Pro Glu Ala Val Ala His Leu Glu Ala Asp Arg Asp Tyr Leu				
	740		745		750
	Glu Ile Glu Val Gly Gly Arg Arg Leu Phe Phe Val Lys Ala His Val				
	755		760		765
	Arg Glu Ser Leu Leu Ser Ile Leu Leu Arg Asp Trp Leu Ala Met Arg				
35	770		775		780
	Lys Gln Ile Arg Ser Arg Ile Pro Gln Ser Thr Pro Glu Glu Ala Val				
	785		790		795 800
	Leu Leu Asp Lys Gln Gln Ala Ala Ile Lys Val Val Cys Asn Ser Val				
	805		810		815
40	Tyr Gly Phe Thr Gly Val Gln His Gly Leu Leu Pro Cys Leu His Val				
	820		825		830
	Ala Ala Thr Val Thr Thr Ile Gly Arg Glu Met Leu Leu Ala Thr Arg				
	835		840		845

Ala Tyr Val His Ala Arg Trp Ala Glu Phe Asp Gln Leu Leu Ala Asp  
 850 855 860  
 Phe Pro Glu Ala Ala Gly Met Arg Ala Pro Gly Pro Tyr Ser Met Arg  
 865 870 875 880  
 5 Ile Ile Tyr Gly Asp Thr Asp Ser Ile Phe Val Leu Cys Arg Gly Leu  
 885 890 895  
 Thr Ala Ala Gly Leu Val Ala Met Gly Asp Lys Met Ala Ser His Arg  
 900 905 910  
 10 Ala Leu Phe Leu Pro Pro Ile Lys Leu Glu Cys Glu Lys Thr Phe Thr  
 915 920 925  
 Lys Leu Leu Leu Ile Ala Lys Lys Lys Tyr Ile Gly Val Ile Cys Gly  
 930 935 940  
 Gly Lys Met Leu Ile Lys Gly Val Asp Leu Val Arg Lys Asn Asn Cys  
 945 950 955 960  
 15 Ala Phe Ile Asp Arg Thr Ser Arg Ala Leu Val Asp Leu Leu Phe Tyr  
 965 970 975  
 Asp Asp Thr Val Ser Gly Ala Ala Ala Ala Glu Arg Pro Ala Glu Glu  
 980 985 990  
 Trp Leu Ala Arg Pro Leu Pro Glu Gly Leu Gln Ala Phe Gly Ala Val  
 995 1000 1005  
 20 Leu Val Asp Ala His Arg Arg Ile Thr Asp Pro Glu Arg Asp Ile Gln  
 1010 1015 1020  
 Asp Phe Val Leu Thr Ala Glu Leu Ser Arg His Pro Arg Ala Tyr Thr  
 1025 1030 1035 1040  
 25 Asn Lys Arg Leu Ala His Leu Thr Val Tyr Tyr Lys Leu Met Ala Arg  
 1045 1050 1055  
 Arg Ala Gln Val Pro Ser Ile Lys Asp Arg Ile Pro Tyr Val Ile Val  
 1060 1065 1070  
 Ala Gln Thr Arg Glu Val Glu Glu Thr Val Ala Arg Leu Ala Ala Leu  
 1075 1080 1085  
 30 Arg Glu Leu Asp Ala Ala Ala Pro Gly Asp Glu Pro Ala Pro Pro Ala  
 1090 1095 1100  
 Ala Leu Pro Ser Pro Ala Lys Arg Pro Arg Glu Thr Pro Ser His Ala  
 1105 1110 1115 1120  
 35 Asp Pro Pro Gly Gly Ala Ser Lys Pro Arg Lys Leu Leu Val Ser Glu  
 1125 1130 1135  
 Leu Ala Glu Asp Pro Gly Tyr Ala Ile Arg Val Pro Leu Asn Thr Asp  
 1140 1145 1150  
 Tyr Tyr Phe Ser His Leu Leu Gly Ala Ala Cys Val Thr Phe Lys Ala  
 1155 1160 1165  
 40 Leu Phe Gly Asn Asn Ala Lys Ile Thr Glu Ser Leu Leu Lys Arg Phe  
 1170 1175 1180  
 Ile Pro Glu Thr Trp His Pro Pro Asp Asp Val Ala Ala Arg Leu Arg  
 621

1185                                      1190                                      1195                                      120  
 Ala Ala Gly Phe Gly Pro Ala Gly Ala Gly Ala Thr Ala Glu Glu Thr  
    1205                                      1210                                      1215  
 Arg Arg Met Leu His Arg Ala Phe Asp Thr Leu Ala  
 5                                      1220                                      1225

## (2) INFORMATION FOR SEQ ID NO:247:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 303 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 15 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

Met Tyr Asp Ile Ala Pro Arg Arg Ser Gly Ser Arg Pro Gly Pro Gly  
 20        1                                      5                                      10                                      15  
 Arg Asp Lys Thr Arg Arg Arg Ser Arg Phe Ser Ala Ala Gly Asn Pro  
    20                                      25                                      30  
 Gly Val Glu Arg Arg Ala Ser Arg Lys Ser Leu Pro Ser His Ala Arg  
    35                                      40                                      45  
 25 Arg Leu Glu Leu Cys Leu His Glu Arg Arg Arg Tyr Arg Gly Phe Phe  
    50                                      55                                      60  
 Ala Ala Gln Thr Pro Ser Glu Glu Ile Ala Ile Val Arg Ser Leu Ser  
 65                                      70                                      75                                      80  
 Val Pro Leu Val Lys Thr Thr Pro Val Ser Leu Pro Phe Ser Leu Asp  
 30                                      85                                      90                                      95  
 Gln Thr Val Ala Asp Asn Cys Leu Thr Leu Ser Gly Met Gly Tyr Tyr  
    100                                      105                                      110  
 Leu Gly Ile Gly Gly Cys Cys Pro Ala Cys Ser Ala Gly Asp Gly Arg  
    115                                      120                                      125  
 35 Leu Ala Thr Val Ser Arg Glu Ala Leu Ile Leu Ala Phe Val Gln Gln  
    130                                      135                                      140  
 Ile Asn Thr Ile Phe Glu His Arg Thr Phe Leu Ala Ser Leu Val Val  
 145                                      150                                      155                                      160  
 Leu Ala Asp Arg His Ser Thr Pro Leu Gln Asp Leu Leu Ala Asp Thr  
 40                                      165                                      170                                      175  
 Leu Gly Gln Pro Glu Leu Phe Phe Val His Thr Ile Leu Arg Gly Gly  
    180                                      185                                      190  
 Gly Ala Cys Asp Pro Arg Phe Leu Phe Tyr Pro Asp Pro Thr Tyr Gly

195                      200                      205  
 Gly His Met Leu Tyr Val Ile Phe Pro Gly Thr Ser Ala His Leu His  
      210                      215                      220  
 Tyr Arg Leu Ile Asp Arg Met Leu Thr Ala Cys Pro Gly Tyr Arg Phe  
 5    225                      230                      235                      240  
 Ala Ala His Val Trp Gln Ser Thr Phe Val Leu Val Val Arg Arg Asn  
                          245                      250                      255  
 Ala Glu Lys Pro Ala Asp Ala Glu Ile Pro Thr Val Ser Ala Ala Asp  
                          260                      265                      270  
 10 Ile Tyr Cys Lys Met Arg Asp Ile Ser Phe Asp Gly Gly Leu Met Leu  
                          275                      280                      285  
 Glu Tyr Gln Arg Leu Tyr Ala Thr Phe Asp Glu Phe Pro Pro Pro  
                          290                      295                      300

15                      (2) INFORMATION FOR SEQ ID NO:248:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 590 amino acids

(B) TYPE: amino acid

20                      (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

25                      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

Met Ala Thr Ser Ala Pro Gly Val Pro Ser Ser Ala Ala Val Arg Glu  
   1                      5                      10                      15  
 Glu Ser Pro Gly Ser Ser Trp Lys Glu Gly Ala Phe Glu Arg Pro Tyr  
 30                      20                      25                      30  
 Val Ala Phe Asp Pro Asp Leu Leu Ala Leu Asn Glu Ala Leu Cys Ala  
                          35                      40                      45  
 Glu Leu Leu Ala Ala Cys His Val Val Gly Val Pro Pro Ala Ser Ala  
                          50                      55                      60  
 35 Leu Asp Glu Asp Val Glu Ser Asp Val Ala Pro Ala Pro Pro Arg Pro  
                          65                      70                      75                      80  
 Arg Gly Ala Ala Arg Glu Ala Ser Gly Gly Arg Gly Pro Gly Ser Arg  
                          85                      90                      95  
 Pro Pro Ala Asp Pro Thr Ala Glu Gly Leu Leu Asp Thr Gly Pro Phe  
 40                      100                      105                      110  
 Ala Ala Ala Ser Val Asp Thr Phe Ala Leu Asp Arg Pro Cys Leu Val  
                          115                      120                      125  
 Cys Arg Thr Ile Glu Leu Tyr Lys Gln Ala Tyr Arg Leu Ser Pro Gln

	130		135		140	
	Trp Val Ala Asp Tyr Ala Phe Leu Cys Ala Lys Cys Leu Gly Ala Pro					
	145		150		155	160
	His Cys Ala Ala Ser Ile Phe Val Ala Ala Phe Glu Phe Val Tyr Val					
5		165		170		175
	Met Asp His His Phe Leu Arg Thr Lys Lys Ala Thr Leu Val Gly Ser					
		180		185		190
	Phe Ala Arg Phe Ala Leu Thr Ile Asn Asp Ile His Arg His Phe Phe					
		195		200		205
10	Leu His Cys Cys Phe Arg Thr Asp Gly Gly Val Pro Gly Arg His Ala					
		210		215		220
	Gln Lys Gln Pro Arg Pro Thr Pro Ser Pro Gly Ala Ala Lys Val Gln					
	225		230		235	240
	Tyr Ser Asn Tyr Ser Phe Leu Ala Gln Ser Ala Thr Arg Ala Leu Ile					
15		245		250		255
	Gly Thr Leu Ala Ser Gly Gly Asp Asp Gly Ala Gly Ala Gly Gly Gly					
		260		265		270
	Ser Gly Thr Gln Pro Ser Leu Thr Thr Ala Leu Met Asn Trp Lys Asp					
		275		280		285
20	Cys Ala Arg Leu Leu Asp Cys Thr Glu Gly Lys Arg Gly Gly Gly Asp					
		290		295		300
	Ser Cys Cys Thr Arg Ala Ala Ala Arg Asn Gly Glu Phe Glu Ala Ala					
	305		310		315	320
	Ala Gly Ala Gln Gly Gly Glu Pro Glu Thr Trp Ala Tyr Ala Asp Leu					
25		325		330		335
	Ile Leu Leu Leu Leu Ala Gly Thr Pro Ala Val Trp Glu Ser Gly Pro					
		340		345		350
	Arg Leu Arg Ala Ala Ala Asp Ala Arg Arg Ala Ala Val Ser Glu Ser					
		355		360		365
30	Trp Glu Ala His Arg Gly Ala Arg Met Arg Asp Ala Ala Pro Arg Phe					
		370		375		380
	Ala Gln Phe Ala Glu Pro Lys Ala Gln Pro Asp Leu Asp Leu Gly Pro					
	385		390		395	400
	Leu Met Ala Thr Val Leu Lys His Gly Arg Gly Arg Gly Arg Thr Gly					
35		405		410		415
	Gly Glu Cys Leu Leu Cys Asn Leu Leu Leu Val Arg Ala Tyr Trp Leu					
		420		425		430
	Ala Met Arg Arg Leu Arg Ala Ser Val Val Arg Tyr Ser Glu Asn Asn					
		435		440		445
40	Thr Ser Leu Phe Asp Cys Ile Val Pro Val Val Asp Gln Leu Glu Ala					
		450		455		460
	Asp Pro Glu Ala Gln Pro Gly Asp Gly Gly Arg Phe Val Ser Leu Leu					
	465		470		475	480

Arg Ala Ala Gly Pro Glu Ala Ile Phe Lys His Met Phe Cys Asp Pro  
                             485                            490                            495  
 Met Cys Ala Ile Thr Glu Met Glu Val Asp Pro Trp Val Leu Phe Gly  
                             500                            505                            510  
 5 His Pro Arg Ala Asp His Arg Asp Glu Leu Gln Leu His Lys Ala Lys  
                             515                            520                            525  
 Leu Ala Cys Gly Asn Glu Phe Glu Gly Arg Val Cys Ile Ala Leu Arg  
                             530                            535                            540  
 Ala Leu Ile Tyr Thr Phe Lys Thr Tyr Gln Val Phe Val Pro Lys Pro  
 10 545                            550                            555                            560  
 Thr Ala Thr Phe Val Arg Glu Ala Gly Ala Leu Leu Arg Arg His Ser  
                             565                            570                            575  
 Ile Ser Leu Leu Ser Leu Glu His Thr Leu Cys Thr Tyr Val  
                             580                            585                            590  
 15

## (2) INFORMATION FOR SEQ ID NO:249:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 128 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

Met Ala Gly Arg Ala Gly Arg Trp Arg Thr Leu Arg Asp Ala Ile Pro  
   1                            5                            10                            15  
 30 Asp Cys Ala Leu Arg Ser Gln Thr Leu Glu Ser Leu Asp Ala Arg Tyr  
                             20                            25                            30  
 Val Ser Arg Asp Gly Ala Gly Asp Ala Ala Val Trp Phe Glu Asp Met  
                             35                            40                            45  
 Thr Pro Ala Glu Leu Glu Val Ile Phe Pro Thr Thr Asp Ala Lys Leu  
 35 50                            55                            60  
 Asn Tyr Leu Ser Arg Thr Gln Arg Leu Ala Ser Leu Leu Thr Tyr Ala  
 65                            70                            75                            80  
 Gly Pro Ile Lys Ala Pro Asp Gly Pro Ala Ala Pro His Thr Gln Asp  
                             85                            90                            95  
 40 Thr Ala Cys Val His Gly Glu Leu Leu Ala Arg Lys Arg Glu Arg Phe  
                             100                            105                            110  
 Ala Ala Val Ile Asn Arg Phe Leu Asp Leu His Gln Ile Leu Arg Gly  
                             115                            120                            125

625



## (2) INFORMATION FOR SEQ ID NO:250:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 112 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 10 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

```

Met Ala Ala Pro Gln Phe His Arg Pro Ser Thr Ile Thr Ala Asp Asn
15 1          5          10          15
Val Arg Ala Leu Gly Met Arg Gly Leu Val Leu Ala Thr Asn Asn Ala
          20          25          30
Gln Phe Ile Met Asp Asn Ser Tyr Pro His Pro His Gly Thr Gln Gly
          35          40          45
20 Ala Val Arg Glu Phe Leu Arg Gly Gln Ala Ala Ala Leu Thr Asp Leu
          50          55          60
Gly Val Thr His Ala Asn Asn Thr Phe Ala Pro Gln Pro Met Phe Ala
65          70          75          80
Gly Asp Ala Ala Ala Glu Trp Leu Arg Pro Ser Phe Gly Leu Lys Arg
25          85          90          95
Thr Tyr Ser Pro Phe Val Val Arg Asp Pro Lys Thr Pro Ser Thr Pro
          100          105          110

```

## (2) INFORMATION FOR SEQ ID NO:251:

30

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 112 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 35 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

40

```

Met Ala Ala Pro Gln Phe His Arg Pro Ser Thr Ile Thr Ala Asp Asn
1          5          10          15
Val Arg Ala Leu Gly Met Arg Gly Leu Val Leu Ala Thr Asn Asn Ala

```

```

                20                25                30
Gln Phe Ile Met Asp Asn Ser Tyr Pro His Pro His Gly Thr Gln Gly
                35                40                45
Ala Val Arg Glu Phe Leu Arg Gly Gln Ala Ala Ala Leu Thr Asp Leu
5      50                55                60
Gly Val Thr His Ala Asn Asn Thr Phe Ala Pro Gln Pro Met Phe Ala
65                70                75                80
Gly Asp Ala Ala Ala Glu Trp Leu Arg Pro Ser Phe Gly Leu Lys Arg
                85                90                95
10    Thr Tyr Ser Pro Phe Val Val Arg Asp Pro Lys Thr Pro Ser Thr Pro
                100                105                110

```

## (2) INFORMATION FOR SEQ ID NO:252:

- 15 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3051 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

20

- (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

```

25 Met Ile Pro Ala Ala Leu Pro His Pro Thr Met Lys Arg Gln Gly Asp
   1                5                10                15
Arg Asp Ile Val Val Thr Gly Val Arg Asn Gln Phe Ala Thr Asp Leu
   20                25                30
Glu Pro Gly Gly Ser Val Ser Cys Met Arg Ser Ser Leu Ser Phe Leu
30      35                40                45
Ser Leu Leu Phe Asp Val Gly Pro Arg Asp Val Leu Ser Ala Glu Ala
   50                55                60
Ile Glu Gly Cys Leu Val Glu Gly Gly Glu Trp Thr Arg Ala Ala Ala
65                70                75                80
35 Gly Ser Gly Pro Pro Arg Met Cys Ser Ile Ile Glu Leu Pro Asn Phe
   85                90                95
Leu Glu Tyr Pro Ala Arg Gly Leu Arg Cys Val Phe Ser Arg Val Tyr
  100                105                110
Gly Glu Val Gly Phe Phe Gly Glu Pro Thr Ala Gly Leu Leu Glu Thr
40      115                120                125
Gln Cys Pro Ala His Thr Phe Phe Ala Gly Pro Trp Ala Met Arg Pro
  130                135                140
Leu Ser Tyr Thr Leu Leu Thr Ile Gly Pro Leu Gly Met Gly Arg Asp

```

	145		150		155		160									
	Gly	Asp	Thr	Ala	Tyr	Leu	Phe	Asp	Pro	His	Gly	Leu	Pro	Ala	Gly	Thr
				165				170						175		
	Pro	Ala	Phe	Ile	Ala	Lys	Val	Arg	Ala	Gly	Asp	Val	Tyr	Pro	Tyr	Leu
5			180					185					190			
	Thr	Tyr	Tyr	Ala	His	Asp	Arg	Pro	Lys	Val	Arg	Trp	Ala	Gly	Ala	Met
			195					200					205			
	Val	Phe	Phe	Val	Pro	Ser	Gly	Pro	Gly	Ala	Val	Ala	Pro	Ala	Asp	Leu
		210					215					220				
10	Thr	Ala	Ala	Ala	Leu	His	Leu	Tyr	Gly	Ala	Ser	Glu	Thr	Tyr	Leu	Gln
		225					230					235				240
	Asp	Glu	Pro	Phe	Val	Glu	Arg	Arg	Val	Ala	Ile	Thr	His	Pro	Leu	Arg
				245						250						255
	Gly	Glu	Ile	Gly	Gly	Leu	Gly	Ala	Leu	Phe	Val	Gly	Val	Val	Pro	Arg
15			260						265					270		
	Gly	Asp	Gly	Glu	Gly	Ser	Gly	Pro	Val	Val	Pro	Ala	Leu	Pro	Ala	Pro
			275					280						285		
	Thr	His	Val	Gln	Thr	Pro	Arg	Ala	Asp	Arg	Pro	Pro	Glu	Ala	Pro	Arg
		290					295						300			
20	Gly	Ala	Ser	Gly	Pro	Pro	Asn	Thr	Pro	Gln	Ala	Gly	His	Pro	Asn	Arg
		305					310					315				320
	Pro	Pro	Asp	Asp	Val	Trp	Ala	Ala	Ala	Leu	Glu	Gly	Thr	Pro	Pro	Ala
				325						330						335
	Lys	Pro	Ser	Ala	Pro	Asp	Ala	Ala	Ala	Ser	Gly	Pro	Pro	His	Ala	Ala
25			340							345				350		
	Pro	Pro	Pro	Gln	Thr	Pro	Ala	Gly	Asp	Ala	Ala	Glu	Glu	Ala	Glu	Asp
			355						360					365		
	Leu	Arg	Val	Leu	Glu	Val	Gly	Ala	Val	Pro	Val	Gly	Arg	His	Arg	Ala
		370					375					380				
30	Arg	Tyr	Ser	Thr	Gly	Leu	Pro	Lys	Arg	Arg	Arg	Pro	Thr	Trp	Thr	Pro
		385					390					395				400
	Pro	Ser	Ser	Val	Glu	Asp	Leu	Thr	Ser	Gly	Glu	Arg	Pro	Ala	Pro	Lys
				405						410						415
	Ala	Pro	Pro	Ala	Lys	Ala	Lys	Lys	Lys	Ser	Ala	Pro	Lys	Lys	Lys	Ala
35			420							425				430		
	Pro	Val	Ala	Ala	Glu	Val	Pro	Ala	Ser	Ser	Pro	Thr	Pro	Ile	Ala	Ala
			435						440					445		
	Thr	Val	Pro	Pro	Ala	Pro	Asp	Thr	Pro	Pro	Gln	Ser	Gly	Gln	Gly	Gly
		450					455					460				
40	Gly	Asp	Asp	Gly	Pro	Asp	Ser	Ser	Pro	Ser	Val	Leu	Glu	Thr	Leu	Gly
		465					470					475				480
	Ala	Arg	Arg	Pro	Pro	Glu	Pro	Pro	Gly	Ala	Asp	Leu	Ala	Gln	Leu	Phe
				485						490				495		

Glu Val His Pro Asn Val Ala Ala Thr Ala Val Arg Leu Ala Ala Arg  
 500 505 510  
 Asp Ala Ala Arg Glu Val Ala Ala Cys Ser Gln Leu Thr Ile Asn Ala  
 515 520 525  
 5 Leu Arg Ser Pro Tyr Pro Ala His Pro Gly Leu Leu Glu Leu Cys Val  
 530 535 540  
 Ile Phe Phe Phe Glu Arg Val Leu Ala Phe Leu Ile Glu Asn Gly Ala  
 545 550 555 560  
 Arg Thr His Thr Gln Ala Gly Val Ala Gly Pro Ala Ala Ala Leu Leu  
 10 565 570 575  
 Asp Phe Thr Leu Arg Met Leu Pro Arg Lys Thr Ala Val Gly Asp Phe  
 580 585 590  
 Leu Ala Ser Thr Arg Met Ser Leu Ala Asp Val Ala Ala His Arg Pro  
 595 600 605  
 15 Leu Ile Gln His Val Leu Asp Glu Asn Ser Gln Ile Gly Arg Leu Ala  
 610 615 620  
 Lys Leu Val Leu Val Ala Arg Asp Val Ile Arg Glu Thr Asp Ala Phe  
 625 630 635 640  
 Tyr Gly Asp Leu Ala Asp Leu Asp Leu Gln Leu Arg Ala Ala Pro Pro  
 20 645 650 655  
 Ala Asn Leu Tyr Ala Arg Leu Gly Glu Trp Leu Leu Glu Arg Ser Arg  
 660 665 670  
 Ala His Pro Asn Thr Leu Phe Ala Pro Ala Thr Pro Thr His Pro Glu  
 675 680 685  
 25 Pro Leu Leu His Arg Ile Gln Ala Gln Phe Arg Glu Glu Met Arg Val  
 690 695 700  
 Glu Ala Glu Ala Arg Glu Met Arg Glu Ala Leu Asp Arg Val Asp Ser  
 705 710 715 720  
 Val Ser Gln Arg Ala Gly Pro Leu Thr Val Met Pro Val Pro Ala Ala  
 30 725 730 735  
 Pro Gly Ala Gly Gly Arg Ala Pro Cys Pro Pro Ala Leu Gly Pro Glu  
 740 745 750  
 Ala Ile Gln Ala Arg Leu Glu Asp Val Arg Ile Gln Ala Arg Arg Ala  
 755 760 765  
 35 Ile Glu Ser Ala Ile Lys Glu Tyr Phe His Arg Gly Ala Val Tyr Ser  
 770 775 780  
 Ala Lys Ala Leu Gln Ala Ser Asp Ser His Asp Cys Arg Phe His Val  
 785 790 795 800  
 Ala Ser Ala Ala Val Val Pro Met Val Gln Leu Leu Glu Ser Leu Pro  
 40 805 810 815  
 Ala Phe Asp Gln His Thr Arg Asp Val Ala Gln Arg Ala Ala Leu Pro  
 820 825 830  
 Pro Pro Pro Pro Leu Ala Thr Ser Pro Gln Ala Ile Leu Leu Arg Asp  
 629

	835	840	845
	Leu Leu Gln Arg Gly Gln Thr Leu Asp Ala Pro Glu Asp Leu Ala Ala		
	850	855	860
	Trp Leu Ser Val Leu Thr Asp Ala Ala Thr Gln Gly Leu Ile Glu Arg		
5	865	870	875 880
	Lys Pro Leu Glu Glu Leu Ala Arg Ser Ile His Gly Ile Asn Asp Gln		
	885	890	895
	Gln Ala Arg Arg Ser Ser Gly Leu Ala Glu Leu Gln Arg Phe Asp Ala		
	900	905	910
10	Leu Asp Ala Ala Gln Gln Leu Asp Ser Asp Ala Ala Phe Val Pro Ala		
	915	920	925
	Thr Gly Pro Ala Pro Tyr Val Asp Gly Gly Gly Leu Ser Pro Glu Ala		
	930	935	940
	Thr Arg Met Ala Glu Asp Ala Leu Arg Gln Ala Arg Ala Met Glu Ala		
15	945	950	955 960
	Ala Lys Met Thr Ala Glu Leu Ala Pro Glu Ala Arg Ser Arg Leu Arg		
	965	970	975
	Glu Arg Ala His Ala Leu Glu Ala Met Leu Asn Asp Ala Arg Glu Arg		
	980	985	990
20	Ala Lys Val Ala His Asp Ala Arg Glu Lys Phe Leu His Lys Leu Gln		
	995	1000	1005
	Gly Val Leu Arg Pro Leu Pro Asp Phe Val Gly Leu Lys Ala Cys Pro		
	1010	1015	1020
	Ala Val Leu Ala Thr Leu Arg Ala Ser Leu Pro Ala Gly Trp Thr Asp		
25	1025	1030	1035 104
	Leu Ala Asp Ala Val Arg Gly Pro Pro Pro Glu Val Thr Ala Ala Leu		
	1045	1050	1055
	Arg Ala Asp Leu Trp Gly Leu Leu Gly Gln Tyr Arg Glu Ala Leu Glu		
	1060	1065	1070
30	His Pro Thr Pro Asp Thr Ala Thr Ala Gly Leu His Pro Ala Phe Val		
	1075	1080	1085
	Val Val Leu Lys Thr Leu Phe Ala Asp Ala Pro Glu Thr Pro Val Leu		
	1090	1095	1100
	Val Gln Phe Phe Ser Asp His Ala Pro Thr Ile Ala Lys Ala Val Ser		
35	1105	1110	1115 112
	Asn Ala Ile Asn Ala Gly Ser Ala Ala Val Ala Thr Asp Ala Ala Thr		
	1125	1130	1135
	Val Asp Ala Ala Val Arg Ala His Gly Ala Asp Ala Val Ser Ala Leu		
	1140	1145	1150
40	Gly Ala Ala Ala Arg Asp Pro Asp Leu Ser Phe Leu Ala Ala Asp Ser		
	1155	1160	1165
	Ala Ala Gly Tyr Val Lys Ala Thr Arg Leu Ala Leu Glu Arg Ala Ile		
	1170	1175	1180

Asp Glu Leu Thr Thr Leu Gly Ser Ala Ala Ala Asp Leu Val Val Gln  
 1185 1190 1195 120  
 Ala Arg Arg Ala Cys Ala Gln Pro Glu Gly Asp His Ala Ala Leu Ile  
 1205 1210 1215  
 5 Asp Ala Ala Ala Arg Ala Thr Thr Ala Ala Arg Glu Ser Leu Ala Gly  
 1220 1225 1230  
 His Glu Ala Gly Phe Gly Gly Leu Leu His Ala Glu Gly Thr Ala Gly  
 1235 1240 1245  
 Asp His Ser Pro Ser Gly Arg Ala Leu Gln Glu Leu Gly Lys Val Ile  
 10 1250 1255 1260  
 Gly Ala Thr Arg Arg Arg Ala Asp Glu Leu Glu Ala Ala Val Ala Asp  
 1265 1270 1275 128  
 Leu Thr Ala Lys Met Ala Ala Gln Arg Arg Ser Ser Trp Ala Ala Gly  
 1285 1290 1295  
 15 Val Glu Ala Ala Leu Asp Arg Val Glu Asn Arg Ala Glu Phe Asp Val  
 1300 1305 1310  
 Val Glu Leu Arg Arg Leu Gln Ala Gly Thr His Gly Tyr Asn Pro Arg  
 1315 1320 1325  
 Asp Phe Arg Lys Arg Ala Glu Gln Ala Ala Asn Ala Glu Ala Val Thr  
 20 1330 1335 1340  
 Leu Ala Leu Asp Thr Ala Phe Ala Phe Asn Pro Tyr Thr Pro Glu Asn  
 1345 1350 1355 136  
 Gln Arg His Pro Met Leu Pro Pro Leu Ala Ala Ile His Arg Leu Gly  
 1365 1370 1375  
 25 Trp Ser Ala Ala Phe His Ala Ala Ala Glu Thr Tyr Ala Asp Met Phe  
 1380 1385 1390  
 Arg Val Asp Ala Glu Pro Leu Ala Arg Leu Leu Arg Ile Ala Glu Gly  
 1395 1400 1405  
 Leu Leu Glu Met Ala Gln Ala Gly Asp Gly Phe Ile Asp Tyr His Glu  
 30 1410 1415 1420  
 Ala Val Gly Arg Leu Ala Asp Asp Met Thr Ser Val Pro Gly Leu Arg  
 1425 1430 1435 144  
 Arg Tyr Val Pro Phe Phe Gln His Gly Tyr Ala Asp Tyr Val Glu Leu  
 1445 1450 1455  
 35 Arg Asp Arg Leu Asp Ala Ile Arg Ala Asp Val His Arg Ala Leu Gly  
 1460 1465 1470  
 Gly Val Pro Leu Asp Leu Ala Ala Ala Ala Glu Gln Ile Ser Ala Ala  
 1475 1480 1485  
 Arg Asn Asp Pro Glu Ala Thr Ala Glu Leu Val Arg Thr Gly Val Thr  
 40 1490 1495 1500  
 Leu Pro Cys Pro Ser Glu Asp Ala Leu Val Ala Cys Ala Ala Ala Leu  
 1505 1510 1515 152  
 Glu Arg Val Asp Gln Ser Pro Val Lys Asn Thr Ala Tyr Ala Glu Tyr

		1525		1530		1535
	Val	Ala	Phe	Val	Thr	Arg
				Gln	Asp	Thr
				Ala	Glu	Thr
				Lys	Asp	Ala
				Val		
		1540		1545		1550
	Val	Arg	Ala	Lys	Gln	Gln
				Arg	Ala	Glu
				Ala	Thr	Glu
				Arg	Val	Met
				Ala		
5		1555		1560		1565
	Gly	Leu	Arg	Glu	Ala	Ala
				Arg	Glu	Arg
				Ala	Gln	Ile
				Glu	Ala	Glu
		1570		1575		1580
	Gly	Leu	Ala	Asn	Leu	Lys
				Thr	Met	Leu
				Lys	Val	Val
				Ala	Val	Pro
				Ala		
	1585		1590		1595	160
10	Thr	Val	Ala	Lys	Thr	Leu
				Asp	Gln	Ala
				Arg	Ser	Val
				Ala	Glu	Ile
				Ala		
		1605		1610		1615
	Asp	Gln	Val	Glu	Val	Leu
				Leu	Asp	Gln
				Thr	Glu	Lys
				Thr	Arg	Glu
				Leu		
		1620		1625		1630
	Asp	Val	Pro	Ala	Val	Ile
				Trp	Leu	Glu
				His	Ala	Gln
				Arg	Thr	Phe
				Glu		
15		1635		1640		1645
	Thr	His	Pro	Leu	Ser	Ala
				Arg	Asp	Gly
				Pro	Gly	Pro
				Leu	Ala	Arg
				His		
		1650		1655		1660
	Ala	Gly	Arg	Leu	Gly	Ala
				Leu	Phe	Asp
				Thr	Arg	Arg
				Arg	Val	Asp
				Ala		
	1665		1670		1675	168
20	Leu	Arg	Arg	Ser	Leu	Glu
				Glu	Ala	Glu
				Ala	Glu	Trp
				Asp	Glu	Val
				Trp		
		1685		1690		1695
	Gly	Arg	Phe	Gly	Arg	Val
				Arg	Gly	Gly
				Ala	Trp	Lys
				Ser	Pro	Glu
				Gly		
		1700		1705		1710
	Phe	Arg	Ala	Met	His	Glu
				Gln	Leu	Arg
				Ala	Leu	Gln
				Asp	Thr	Thr
				Asn		
25		1715		1720		1725
	Thr	Val	Ser	Gly	Leu	Arg
				Ala	Gln	Pro
				Ala	Tyr	Glu
				Arg	Leu	Ser
				Ala		
		1730		1735		1740
	Arg	Tyr	Gln	Gly	Val	Leu
				Gly	Ala	Lys
				Gly	Ala	Glu
				Arg	Ala	Glu
				Ala		
	1745		1750		1755	176
30	Val	Glu	Glu	Leu	Gly	Ala
				Arg	Val	Thr
				Lys	His	Thr
				Ala	Leu	Cys
				Ala		
		1765		1770		1775
	Arg	Leu	Arg	Asp	Glu	Val
				Val	Arg	Arg
				Val	Pro	Trp
				Glu	Met	Asn
				Phe		
		1780		1785		1790
	Asp	Ala	Leu	Gly	Arg	Leu
				Leu	Ala	Glu
				Phe	Asp	Ala
				Ala	Ala	Ala
				Asp		
35		1795		1800		1805
	Leu	Ala	Pro	Trp	Ala	Val
				Glu	Glu	Phe
				Arg	Gly	Ala
				Arg	Glu	Leu
				Ile		
		1810		1815		1820
	Gln	Arg	Arg	Met	Gly	Ser
				Ala	Tyr	Ala
				Arg	Ala	Gly
				Gly	Gly	Gln
				Thr	Gly	
	1825		1830		1835	184
40	Ala	Gly	Ala	Ala	Ala	Pro
				Ala	Pro	Leu
				Val	Asp	Leu
				Arg		
		1845		1850		1855
	Leu	Asp	Ala	Arg	Ala	Arg
				Ala	Ser	Ser
				Ser	Pro	Glu
				Gly	His	Glu
				Val		
		1860		1865		1870

Asp Pro Gln Leu Leu Arg Arg Arg Gly Glu Ala Tyr Leu Arg Ala Gly  
 1875 1880 1885  
 Gly Asp Pro Gly Pro Leu Val Leu Arg Glu Ala Val Ser Ala Leu Asp  
 1890 1895 1900  
 5 Leu Pro Phe Ala Thr Ser Phe Leu Ala Pro Asp Gly Thr Pro Leu Gln  
 1905 1910 1915 192  
 Tyr Ala Leu Cys Phe Pro Ala Val Thr Asp Lys Leu Gly Ala Leu Leu  
 1925 1930 1935  
 Met Arg Pro Glu Ala Ala Cys Val Arg Pro Pro Leu Pro Thr Asp Val  
 10 1940 1945 1950  
 Leu Glu Ser Ala Pro Thr Val Thr Ala Met Tyr Val Leu Thr Val Val  
 1955 1960 1965  
 Asn Arg Leu Gln Leu Ala Leu Ser Asp Ala Gln Ala Ala Asn Phe Gln  
 1970 1975 1980  
 15 Leu Phe Gly Arg Phe Val Arg His Arg Gln Ala Thr Trp Gly Ala Ser  
 1985 1990 1995 200  
 Met Asp Ala Ala Ala Glu Leu Tyr Val Val Ala Thr Thr Leu Thr Arg  
 2005 2010 2015  
 Glu Phe Gly Cys Arg Trp Ala Gln Leu Gly Trp Ala Ser Gly Ala Ala  
 20 2020 2025 2030  
 Ala Pro Arg Pro Pro Pro Gly Pro Arg Gly Ser Gln Arg His Cys Val  
 2035 2040 2045  
 Ala Phe Asn Glu Asn Asp Val Leu Val Val Ala Gly Val Pro Glu His  
 2050 2055 2060  
 25 Ile Tyr Asn Phe Trp Arg Leu Asp Leu Val Arg Gln His Glu Tyr Met  
 2065 2070 2075 208  
 His Leu Thr Leu Glu Arg Ala Phe Glu Asp Ala Ala Glu Ser Met Leu  
 2085 2090 2095  
 Phe Val Gln Arg Leu Thr Pro His Pro Asp Ala Arg Ile Arg Val Leu  
 30 2100 2105 2110  
 Pro Thr Phe Leu Asp Gly Gly Pro Pro Thr Arg Gly Leu Leu Phe Gly  
 2115 2120 2125  
 Thr Arg Leu Ala Asp Trp Arg Arg Gly Lys Leu Ser Glu Thr Asp Pro  
 2130 2135 2140  
 35 Leu Ala Pro Trp Arg Ser Ala Leu Glu Leu Gly Thr Gln Arg Arg Asp  
 2145 2150 2155 216  
 Ala Pro Ala Leu Gly Lys Leu Ser Pro Ala Gln Ala Ala Val Ser Val  
 2165 2170 2175  
 Leu Gly Arg Met Cys Leu Pro Ser Ala Ala Ala Leu Trp Thr Cys Met  
 40 2180 2185 2190  
 Phe Pro Asp Asp Tyr Thr Glu Tyr Asp Ser Phe Asp Ala Leu Leu Ala  
 2195 2200 2205  
 Ala Arg Leu Glu Ser Gly Gln Thr Leu Gly Pro Ala Gly Gly Arg Glu



	2210	2215	2220
	Ala Ser Leu Pro Glu Ala Pro His Ala Leu Tyr Arg Pro Thr Gly Gln		
	2225	2230	2235 224
	His Val Ala Val Leu Ala Ala Ala Thr Thr Pro Ala Ala Arg Val Thr		
5	2245	2250	2255
	Ala Met Asp Leu Val Leu Ala Ala Val Leu Leu Gly Ala Pro Val Val		
	2260	2265	2270
	Val Arg Asn Thr Thr Ala Phe Ser Arg Glu Ser Glu Leu Glu Leu Cys		
	2275	2280	2285
10	Leu Thr Leu Phe Asp Ser Arg Pro Gly Gly Pro Asp Ala Ala Leu Arg		
	2290	2295	2300
	Asp Val Val Ser Ser Asp Ile Glu Thr Trp Ala Val Gly Leu Leu His		
	2305	2310	2315 232
	Thr Asp Leu Asn Pro Ile Glu Asn Ala Cys Leu Ala Ala Gln Leu Pro		
15	2325	2330	2335
	Arg Leu Ser Ala Leu Ile Ala Glu Arg Pro Leu Ala Asp Gly Pro Pro		
	2340	2345	2350
	Cys Leu Val Leu Val Asp Ile Ser Met Thr Pro Val Ala Val Leu Trp		
	2355	2360	2365
20	Glu Ala Pro Glu Pro Pro Gly Pro Pro Asp Val Arg Phe Val Gly Ser		
	2370	2375	2380
	Glu Ala Thr Glu Glu Leu Pro Phe Val Ala Thr Ala Gly Asp Val Leu		
	2385	2390	2395 240
	Ala Ala Ser Ala Ala Asp Ala Asp Pro Phe Phe Ala Arg Ala Ile Leu		
25	2405	2410	2415
	Gly Arg Pro Phe Asp Ala Ser Leu Leu Thr Gly Glu Leu Phe Pro Gly		
	2420	2425	2430
	His Pro Val Tyr Gln Arg Pro Leu Ala Asp Glu Ala Gly Pro Ser Ala		
	2435	2440	2445
30	Pro Thr Ala Ala Arg Asp Pro Arg Asp Leu Ala Gly Gly Asp Gly Gly		
	2450	2455	2460
	Ser Gly Pro Glu Asp Pro Ala Ala Pro Pro Ala Arg Gln Ala Asp Pro		
	2465	2470	2475 248
	Gly Val Leu Ala Pro Thr Leu Leu Thr Asp Ala Thr Thr Gly Glu Pro		
35	2485	2490	2495
	Val Pro Pro Arg Met Trp Ala Trp Ile His Gly Leu Glu Glu Leu Ala		
	2500	2505	2510
	Ser Asp Asp Ala Gly Gly Pro Thr Pro Asn Pro Ala Pro Ala Leu Leu		
	2515	2520	2525
40	Pro Pro Pro Ala Thr Asp Gln Ser Val Pro Thr Ser Gln Tyr Ala Pro		
	2530	2535	2540
	Arg Pro Ile Gly Pro Ala Ala Thr Ala Arg Glu Trp Ser Val Pro Pro		
	2545	2550	2555 256

Gln Gln Asn Thr Gly Arg Val Pro Val Ala Pro Arg Asp Asp Pro Arg  
 2565 2570 2575  
 Pro Ser Pro Pro Thr Pro Ser Pro Pro Ala Asp Ala Ala Leu Pro Pro  
 2580 2585 2590  
 5 Pro Ala Phe Ser Gly Ser Ala Ala Ala Phe Ser Ala Ala Val Pro Arg  
 2595 2600 2605  
 Val Arg Arg Ser Arg Arg Thr Arg Ala Lys Ser Arg Ala Pro Arg Ala  
 2610 2615 2620  
 Ser Ala Pro Pro Glu Gly Trp Arg Pro Pro Ala Leu Pro Ala Pro Val  
 10 2625 2630 2635 264  
 Ala Pro Val Ala Ala Ser Ala Arg Pro Pro Asp Gln Pro Pro Thr Pro  
 2645 2650 2655  
 Glu Ser Ala Pro Pro Ala Trp Val Ser Ala Leu Pro Leu Pro Pro Gly  
 2660 2665 2670  
 15 Pro Ala Ser Arg Ala Phe Pro Ala Pro Thr Leu Ala Pro Ile Pro Pro  
 2675 2680 2685  
 Pro Pro Ala Glu Gly Ala Val Ala Pro Gly Asp Asp Arg Arg Arg Gly  
 2690 2695 2700  
 Arg Arg Gln Thr Thr Ala Gly Pro Ser Pro Thr Pro Pro Arg Gly Pro  
 20 2705 2710 2715 272  
 Ala Ala Gly Pro Pro Arg Arg Leu Trp Ala Val Ala Ser Leu Ser Ala  
 2725 2730 2735  
 Ser Leu Asn Ser Leu Pro Ser Pro Arg Asp Pro Ala Asp His Ala Ala  
 2740 2745 2750  
 25 Ala Val Ser Ala Ala Ala Ala Val Pro Pro Ser Pro Gly Leu Ala  
 2755 2760 2765  
 Pro Pro Thr Ser Ala Val Gln Thr Ser Pro Pro Pro Leu Ala Pro Gly  
 2770 2775 2780  
 Pro Val Ala Pro Ser Glu Pro Leu Cys Gly Trp Val Val Pro Gly Gly  
 30 2785 2790 2795 280  
 Pro Val Ala Arg Arg Pro Pro Pro Gln Ser Pro Ala Thr Lys Pro Ala  
 2805 2810 2815  
 Ala Arg Thr Arg Ile Arg Ala Arg Ser Val Pro Gln Pro Pro Leu Pro  
 2820 2825 2830  
 35 Gln Pro Pro Leu Pro Gln Pro Pro Leu Pro Gln Pro Pro Leu Pro Gln  
 2835 2840 2845  
 Pro Pro Leu Pro Gln Pro Pro Leu Pro Gln Pro Pro Leu Pro Gln Pro  
 2850 2855 2860  
 Pro Leu Pro Gln Pro Pro Leu Pro Gln Pro Pro Leu Pro Gln Pro Pro  
 40 2865 2870 2875 288  
 Leu Pro Gln Pro Pro Leu Pro Gln Ser Arg Asp Ser Val Pro Thr Pro  
 2885 2890 2895  
 Glu Ser Pro Thr His Thr Asn Thr His Leu Pro Val Ser Ala Val Thr

[illegible]

(2) INFORMATION FOR SEQ ID NO:253:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1124 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

[illegible]

		85		90		95	
	Ala Gly Gly Asp Gly Arg Thr Ala Val Val Gly Gly Thr Ser Gly Pro						
		100		105		110	
	Ser Ala Thr Thr Ser Val Gly Thr Gln Thr Ser Gly Glu Phe Leu His						
5		115		120		125	
	Gly Asn Pro Arg Thr Pro Glu Pro Gln Gly Pro Gln Ala Val Pro Pro						
		130		135		140	
	Pro Pro Pro Pro Pro Phe Pro Trp Gly His Glu Cys Cys Ala Arg Arg						
	145		150		155		160
10	Asp Arg Gly Ala Glu Lys Asp Val Gly Ala Ala Glu Ser Trp Ser Asp						
		165		170		175	
	Gly Pro Ser Ser Asp Ser Glu Thr Glu Asp Ser Asp Ser Ser Asp Glu						
		180		185		190	
	Asp Thr Gly Ser Gly Ser Glu Thr Leu Ser Arg Ser Ser Ser Ile Trp						
15		195		200		205	
	Ala Ala Gly Ala Thr Asp Asp Asp Asp Ser Asp Ser Asp Ser Arg Ser						
		210		215		220	
	Asp Asp Ser Val Gln Pro Asp Val Val Val Arg Arg Arg Trp Ser Asp						
	225		230		235		240
20	Gly Pro Ala Pro Val Ala Phe Pro Lys Pro Arg Arg Pro Gly Asp Ser						
		245		250		255	
	Pro Gly Asn Pro Gly Leu Gly Ala Gly Thr Gly Pro Gly Ser Ala Thr						
		260		265		270	
	Asp Pro Arg Ala Ser Ala Asp Ser Asp Ser Ala Ala His Ala Ala Ala						
25		275		280		285	
	Pro Gln Ala Asp Val Ala Pro Val Leu Asp Ser Gln Pro Thr Val Gly						
		290		295		300	
	Thr Asp Pro Gly Tyr Pro Val Pro Leu Glu Leu Thr Pro Glu Asn Ala						
	305		310		315		320
30	Glu Ala Val Ala Arg Phe Leu Gly Asp Ala Val Asp Arg Glu Pro Ala						
		325		330		335	
	Leu Met Leu Glu Tyr Phe Cys Arg Cys Ala Arg Glu Glu Ser Lys Arg						
		340		345		350	
	Val Pro Pro Arg Thr Phe Gly Ser Ala Pro Arg Leu Thr Glu Asp Asp						
35		355		360		365	
	Phe Gly Leu Leu Asn Tyr Ala Glu Met Arg Arg Leu Cys Leu Asp Leu						
		370		375		380	
	Pro Pro Val Pro Pro Asn Ala Tyr Thr Pro Tyr His Leu Arg Glu Tyr						
	385		390		395		400
40	Ala Thr Arg Leu Val Asn Gly Phe Lys Pro Leu Val Arg Arg Ser Ala						
		405		410		415	
	Arg Leu Tyr Arg Ile Leu Gly Ile Leu Val His Leu Arg Ile Arg Thr						
		420		425		430	

Arg Glu Ala Ser Phe Glu Glu Trp Met Arg Ser Lys Glu Val Asp Leu  
 435 440 445  
 Asp Phe Gly Leu Thr Glu Arg Leu Arg Glu His Glu Ala Gln Leu Met  
 450 455 460  
 5 Ile Leu Ala Gln Ala Leu Asn Pro Tyr Asp Cys Leu Ile His Ser Thr  
 465 470 475 480  
 Pro Asn Thr Leu Val Glu Arg Gly Leu Gln Ser Ala Leu Lys Tyr Glu  
 485 490 495  
 10 Glu Phe Tyr Leu Lys Arg Phe Gly Gly His Tyr Met Glu Ser Val Phe  
 500 505 510  
 Gln Met Tyr Thr Arg Ile Ala Gly Phe Leu Ala Cys Arg Ala Thr Arg  
 515 520 525  
 Gly Met Arg His Ile Ala Leu Gly Arg Gln Gly Ser Trp Trp Glu Met  
 530 535 540  
 15 Phe Lys Phe Phe Phe His Arg Leu Tyr Asp His Gln Ile Val Pro Ser  
 545 550 555 560  
 Thr Pro Ala Met Leu Asn Leu Gly Thr Arg Asn Tyr Tyr Thr Ser Ser  
 565 570 575  
 Cys Tyr Leu Val Asn Pro Gln Ala Thr Thr Asn Gln Ala Thr Leu Arg  
 20 580 585 590  
 Ala Ile Thr Gly Asn Val Ser Ala Ile Leu Ala Arg Asn Gly Gly Ile  
 595 600 605  
 Gly Leu Cys Met Gln Ala Phe Asn Asp Asp Gly Thr Ala Ser Ile Met  
 610 615 620  
 25 Pro Ala Leu Lys Val Leu Asp Ser Leu Val Ala Ala His Asn Lys Gln  
 625 630 635 640  
 Ser Trp Thr Gly Ala Cys Val Tyr Leu Glu Pro Trp His Ser Asp Val  
 645 650 655  
 Arg Ala Val Leu Arg Met Lys Gly Val Leu Ala Gly Glu Glu Ala Gln  
 30 660 665 670  
 Arg Cys Asp Asn Ile Phe Ser Ala Leu Trp Met Pro Asp Leu Phe Phe  
 675 680 685  
 Lys Arg Leu Ile Arg His Leu Asp Gly Glu Lys Asn Val Thr Trp Ser  
 690 695 700  
 35 Leu Phe Asp Arg Asp Thr Ser Met Ser Leu Ala Asp Phe His Gly Glu  
 705 710 715 720  
 Glu Phe Glu Lys Leu Tyr Glu His Leu Glu Ala Met Gly Phe Gly Glu  
 725 730 735  
 Thr Ile Pro Ile Gln Asp Leu Ala Tyr Ala Ile Val Arg Ser Ala Ala  
 40 740 745 750  
 Thr Thr Gly Ser Pro Phe Ile Met Phe Lys Asp Ala Val Asn Arg His  
 755 760 765  
 Tyr Ile Tyr Asp Thr Gln Gly Ala Ala Ile Ala Gly Ser Asn Leu Cys  
 638

	770		775		780	
	Thr Glu Ile Val His Pro Ser Ser Lys Arg Ser Ser Gly Val Cys Asn					
	785		790		795	800
	Leu Gly Ser Val Asn Leu Ala Arg Cys Val Ser Arg Arg Thr Phe Asp					
5		805		810		815
	Phe Gly Met Leu Arg Asp Ala Val Gln Ala Cys Val Leu Met Val Asn					
		820		825		830
	Ile Met Ile Asp Ser Thr Leu Gln Pro Thr Pro Gln Cys Arg His Asp					
		835		840		845
10	Asn Leu Arg Ser Met Gly Ile Gly Met Gln Gly Leu His Thr Ala Cys					
		850		855		860
	Leu Lys Met Gly Leu Asp Leu Glu Ser Ala Glu Phe Arg Asp Leu Asn					
		865		870		875
	Thr His Ile Ala Glu Val Met Leu Leu Ala Ala Met Lys Thr Ser Asn					
15		885		890		895
	Ala Leu Cys Val Arg Gly Ala Arg Pro Phe Ser His Phe Lys Arg Ser					
		900		905		910
	Met Tyr Arg Ala Gly Arg Phe His Trp Glu Arg Phe Ser Asn Asp Arg					
		915		920		925
20	Tyr Glu Gly Glu Trp Glu Met Leu Arg Gln Ser Met Met Lys His Gly					
		930		935		940
	Leu Arg Asn Ser Gln Phe Ile Ala Leu Met Pro Thr Ala Ala Ser Ala					
		945		950		955
	Gln Ile Ser Asp Val Ser Glu Gly Phe Ala Pro Leu Phe Thr Asn Leu					
25		965		970		975
	Phe Ser Lys Val Thr Arg Asp Gly Glu Thr Leu Arg Pro Asn Thr Leu					
		980		985		990
	Leu Leu Lys Glu Leu Glu Arg Thr Phe Gly Gly Lys Arg Leu Leu Asp					
		995		1000		1005
30	Ala Met Asp Gly Leu Glu Ala Lys Gln Trp Ser Val Ala Gln Ala Leu					
		1010		1015		1020
	Pro Cys Leu Asp Pro Ala His Pro Leu Arg Arg Phe Lys Thr Ala Phe					
		1025		1030		1035
	Asp Tyr Asp Gln Glu Leu Leu Ile Asp Leu Cys Ala Asp Arg Ala Pro					
35		1045		1050		1055
	Tyr Val Asp His Ser Gln Ser Met Thr Leu Tyr Val Thr Glu Lys Ala					
		1060		1065		1070
	Asp Gly Thr Leu Pro Ala Ser Thr Leu Val Arg Leu Leu Val His Ala					
		1075		1080		1085
40	Tyr Lys Arg Gly Leu Lys Thr Gly Met Tyr Tyr Cys Lys Val Arg Lys					
		1090		1095		1100
	Ala Thr Asn Ser Gly Val Phe Ala Gly Asp Asp Asn Ile Val Cys Thr					
		1105		1110		1115
						112

Ser Cys Ala Leu

(2) INFORMATION FOR SEQ ID NO:254:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1124 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

10

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

15

	Met	Ala	Asn	Arg	Pro	Ala	Ala	Ser	Ala	Gly	Ala	Arg	Ser	Pro	Ser	Gln
	1				5					10					15	
	Glu	Pro	Arg	Glu	Pro	Glu	Val	Ala	Pro	Pro	Gly	Gly	Asp	His	Val	Phe
				20					25					30		
20	Cys	Arg	Lys	Val	Ser	Gly	Val	Met	Val	Leu	Ser	Ser	Asp	Pro	Pro	Gly
			35					40					45			
	Pro	Ala	Ala	Tyr	Arg	Ile	Ser	Asp	Ser	Ser	Phe	Val	Gln	Cys	Gly	Ser
		50				55					60					
	Asn	Cys	Ser	Met	Ile	Ile	Asp	Gly	Asp	Val	Arg	His	Leu	Arg	Asp	Leu
25	65				70					75				80		
	Glu	Gly	Ala	Thr	Ser	Thr	Gly	Ala	Phe	Val	Ala	Ile	Ser	Asn	Val	Ala
				85					90					95		
	Ala	Gly	Gly	Asp	Gly	Arg	Thr	Ala	Val	Val	Gly	Gly	Thr	Ser	Gly	Pro
			100					105					110			
30	Ser	Ala	Thr	Thr	Ser	Val	Gly	Thr	Gln	Thr	Ser	Gly	Glu	Phe	Leu	His
			115					120					125			
	Gly	Asn	Pro	Arg	Thr	Pro	Glu	Pro	Gln	Gly	Pro	Gln	Ala	Val	Pro	Pro
		130				135					140					
	Pro	Pro	Pro	Pro	Pro	Phe	Pro	Trp	Gly	His	Glu	Cys	Cys	Ala	Arg	Arg
35	145				150					155				160		
	Asp	Arg	Gly	Ala	Glu	Lys	Asp	Val	Gly	Ala	Ala	Glu	Ser	Trp	Ser	Asp
				165					170					175		
	Gly	Pro	Ser	Ser	Asp	Ser	Glu	Thr	Glu	Asp	Ser	Asp	Ser	Ser	Asp	Glu
			180					185					190			
40	Asp	Thr	Gly	Ser	Gly	Ser	Glu	Thr	Leu	Ser	Arg	Ser	Ser	Ser	Ile	Trp
			195					200					205			
	Ala	Ala	Gly	Ala	Thr	Asp	Asp	Asp	Asp	Ser	Asp	Ser	Asp	Ser	Arg	Ser
		210					215						220			

Asp Asp Ser Val Gln Pro Asp Val Val Val Arg Arg Arg Trp Ser Asp  
 225 230 235 240  
 Gly Pro Ala Pro Val Ala Phe Pro Lys Pro Arg Arg Pro Gly Asp Ser  
 245 250 255  
 5 Pro Gly Asn Pro Gly Leu Gly Ala Gly Thr Gly Pro Gly Ser Ala Thr  
 260 265 270  
 Asp Pro Arg Ala Ser Ala Asp Ser Asp Ser Ala Ala His Ala Ala Ala  
 275 280 285  
 Pro Gln Ala Asp Val Ala Pro Val Leu Asp Ser Gln Pro Thr Val Gly  
 10 290 295 300  
 Thr Asp Pro Gly Tyr Pro Val Pro Leu Glu Leu Thr Pro Glu Asn Ala  
 305 310 315 320  
 Glu Ala Val Ala Arg Phe Leu Gly Asp Ala Val Asp Arg Glu Pro Ala  
 325 330 335  
 15 Leu Met Leu Glu Tyr Phe Cys Arg Cys Ala Arg Glu Glu Ser Lys Arg  
 340 345 350  
 Val Pro Pro Arg Thr Phe Gly Ser Ala Pro Arg Leu Thr Glu Asp Asp  
 355 360 365  
 Phe Gly Leu Leu Asn Tyr Ala Glu Met Arg Arg Leu Cys Leu Asp Leu  
 20 370 375 380  
 Pro Pro Val Pro Pro Asn Ala Tyr Thr Pro Tyr His Leu Arg Glu Tyr  
 385 390 395 400  
 Ala Thr Arg Leu Val Asn Gly Phe Lys Pro Leu Val Arg Arg Ser Ala  
 405 410 415  
 25 Arg Leu Tyr Arg Ile Leu Gly Ile Leu Val His Leu Arg Ile Arg Thr  
 420 425 430  
 Arg Glu Ala Ser Phe Glu Glu Trp Met Arg Ser Lys Glu Val Asp Leu  
 435 440 445  
 Asp Phe Gly Leu Thr Glu Arg Leu Arg Glu His Glu Ala Gln Leu Met  
 30 450 455 460  
 Ile Leu Ala Gln Ala Leu Asn Pro Tyr Asp Cys Leu Ile His Ser Thr  
 465 470 475 480  
 Pro Asn Thr Leu Val Glu Arg Gly Leu Gln Ser Ala Leu Lys Tyr Glu  
 485 490 495  
 35 Glu Phe Tyr Leu Lys Arg Phe Gly Gly His Tyr Met Glu Ser Val Phe  
 500 505 510  
 Gln Met Tyr Thr Arg Ile Ala Gly Phe Leu Ala Cys Arg Ala Thr Arg  
 515 520 525  
 Gly Met Arg His Ile Ala Leu Gly Arg Gln Gly Ser Trp Trp Glu Met  
 40 530 535 540  
 Phe Lys Phe Phe Phe His Arg Leu Tyr Asp His Gln Ile Val Pro Ser  
 545 550 555 560  
 Thr Pro Ala Met Leu Asn Leu Gly Thr Arg Asn Tyr Tyr Thr Ser Ser



		565		570		575										
	Cys	Tyr	Leu	Val	Asn	Pro	Gln	Ala	Thr	Thr	Asn	Gln	Ala	Thr	Leu	Arg
				580					585					590		
	Ala	Ile	Thr	Gly	Asn	Val	Ser	Ala	Ile	Leu	Ala	Arg	Asn	Gly	Gly	Ile
5			595					600					605			
	Gly	Leu	Cys	Met	Gln	Ala	Phe	Asn	Asp	Asp	Gly	Thr	Ala	Ser	Ile	Met
		610				615					620					
	Pro	Ala	Leu	Lys	Val	Leu	Asp	Ser	Leu	Val	Ala	Ala	His	Asn	Lys	Gln
	625					630					635				640	
10	Ser	Trp	Thr	Gly	Ala	Cys	Val	Tyr	Leu	Glu	Pro	Trp	His	Ser	Asp	Val
				645						650				655		
	Arg	Ala	Val	Leu	Arg	Met	Lys	Gly	Val	Leu	Ala	Gly	Glu	Glu	Ala	Gln
			660					665					670			
	Arg	Cys	Asp	Asn	Ile	Phe	Ser	Ala	Leu	Trp	Met	Pro	Asp	Leu	Phe	Phe
15			675					680					685			
	Lys	Arg	Leu	Ile	Arg	His	Leu	Asp	Gly	Glu	Lys	Asn	Val	Thr	Trp	Ser
		690					695					700				
	Leu	Phe	Asp	Arg	Asp	Thr	Ser	Met	Ser	Leu	Ala	Asp	Phe	His	Gly	Glu
	705				710					715					720	
20	Glu	Phe	Glu	Lys	Leu	Tyr	Glu	His	Leu	Glu	Ala	Met	Gly	Phe	Gly	Glu
				725						730				735		
	Thr	Ile	Pro	Ile	Gln	Asp	Leu	Ala	Tyr	Ala	Ile	Val	Arg	Ser	Ala	Ala
			740						745				750			
	Thr	Thr	Gly	Ser	Pro	Phe	Ile	Met	Phe	Lys	Asp	Ala	Val	Asn	Arg	His
25			755					760					765			
	Tyr	Ile	Tyr	Asp	Thr	Gln	Gly	Ala	Ala	Ile	Ala	Gly	Ser	Asn	Leu	Cys
		770					775					780				
	Thr	Glu	Ile	Val	His	Pro	Ser	Ser	Lys	Arg	Ser	Ser	Gly	Val	Cys	Asn
	785				790					795					800	
30	Leu	Gly	Ser	Val	Asn	Leu	Ala	Arg	Cys	Val	Ser	Arg	Arg	Thr	Phe	Asp
				805						810				815		
	Phe	Gly	Met	Leu	Arg	Asp	Ala	Val	Gln	Ala	Cys	Val	Leu	Met	Val	Asn
			820						825				830			
	Ile	Met	Ile	Asp	Ser	Thr	Leu	Gln	Pro	Thr	Pro	Gln	Cys	Arg	His	Asp
35			835					840					845			
	Asn	Leu	Arg	Ser	Met	Gly	Ile	Gly	Met	Gln	Gly	Leu	His	Thr	Ala	Cys
		850					855					860				
	Leu	Lys	Met	Gly	Leu	Asp	Leu	Glu	Ser	Ala	Glu	Phe	Arg	Asp	Leu	Asn
	865				870					875				880		
40	Thr	His	Ile	Ala	Glu	Val	Met	Leu	Leu	Ala	Ala	Met	Lys	Thr	Ser	Asn
				885						890				895		
	Ala	Leu	Cys	Val	Arg	Gly	Ala	Arg	Pro	Phe	Ser	His	Phe	Lys	Arg	Ser
			900						905				910			

Met Tyr Arg Ala Gly Arg Phe His Trp Glu Arg Phe Ser Asn Asp Arg  
           915                          920                          925  
 Tyr Glu Gly Glu Trp Glu Met Leu Arg Gln Ser Met Met Lys His Gly  
           930                          935                          940  
 5 Leu Arg Asn Ser Gln Phe Ile Ala Leu Met Pro Thr Ala Ala Ser Ala  
    945                          950                          955                          960  
 Gln Ile Ser Asp Val Ser Glu Gly Phe Ala Pro Leu Phe Thr Asn Leu  
                           965                          970                          975  
 Phe Ser Lys Val Thr Arg Asp Gly Glu Thr Leu Arg Pro Asn Thr Leu  
 10                          980                          985                          990  
 Leu Leu Lys Glu Leu Glu Arg Thr Phe Gly Gly Lys Arg Leu Leu Asp  
           995                          1000                          1005  
 Ala Met Asp Gly Leu Glu Ala Lys Gln Trp Ser Val Ala Gln Ala Leu  
           1010                          1015                          1020  
 15 Pro Cys Leu Asp Pro Ala His Pro Leu Arg Arg Phe Lys Thr Ala Phe  
    1025                          1030                          1035                          104  
 Asp Tyr Asp Gln Glu Leu Leu Ile Asp Leu Cys Ala Asp Arg Ala Pro  
                           1045                          1050                          1055  
 Tyr Val Asp His Ser Gln Ser Met Thr Leu Tyr Val Thr Glu Lys Ala  
 20                          1060                          1065                          1070  
 Asp Gly Thr Leu Pro Ala Ser Thr Leu Val Arg Leu Leu Val His Ala  
           1075                          1080                          1085  
 Tyr Lys Arg Gly Leu Lys Thr Gly Met Tyr Tyr Cys Lys Val Arg Lys  
           1090                          1095                          1100  
 25 Ala Thr Asn Ser Gly Val Phe Ala Gly Asp Asp Asn Ile Val Cys Thr  
    1105                          1110                          1115                          112  
 Ser Cys Ala Leu

30 (2) INFORMATION FOR SEQ ID NO:255:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 333 amino acids

(B) TYPE: amino acid

35 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:

Met Asp Pro Ala Val Ser Pro Ala Ser Thr Asp Pro Leu Asp Thr His  
   1                          5                          10                          15

	Ala	Ser	Gly	Ala	Gly	Ala	Ala	Pro	Ile	Pro	Val	Cys	Pro	Thr	Pro	Glu
				20					25					30		
	Arg	Tyr	Phe	Tyr	Thr	Ser	Gln	Cys	Pro	Asp	Ile	Asn	His	Leu	Arg	Ser
			35					40					45			
5	Leu	Ser	Ile	Leu	Asn	Arg	Trp	Leu	Glu	Thr	Glu	Leu	Val	Phe	Val	Gly
		50					55					60				
	Asp	Glu	Glu	Asp	Val	Ser	Lys	Leu	Ser	Glu	Gly	Glu	Leu	Gly	Phe	Tyr
	65					70					75				80	
	Arg	Phe	Leu	Phe	Ala	Phe	Leu	Ser	Ala	Ala	Asp	Asp	Leu	Val	Thr	Glu
10					85					90					95	
	Asn	Leu	Gly	Gly	Leu	Ser	Gly	Leu	Phe	Glu	Gln	Lys	Asp	Ile	Leu	His
				100					105					110		
	Tyr	Tyr	Val	Glu	Gln	Glu	Cys	Ile	Glu	Val	Val	His	Ser	Arg	Val	Tyr
			115					120					125			
15	Asn	Ile	Ile	Gln	Leu	Val	Leu	Phe	His	Asn	Asn	Asp	Gln	Ala	Arg	Arg
			130					135					140			
	Ala	Tyr	Val	Ala	Arg	Thr	Ile	Asn	His	Pro	Ala	Ile	Arg	Val	Lys	Val
	145					150					155				160	
	Asp	Trp	Leu	Glu	Ala	Arg	Val	Arg	Glu	Cys	Asp	Ser	Ile	Pro	Glu	Lys
20					165					170					175	
	Phe	Ile	Leu	Met	Ile	Leu	Ile	Glu	Gly	Val	Phe	Phe	Ala	Ala	Ser	Phe
				180					185					190		
	Ala	Ala	Ile	Ala	Tyr	Leu	Arg	Thr	Asn	Asn	Leu	Leu	Arg	Val	Thr	Cys
			195					200					205			
25	Gln	Ser	Asn	Asp	Leu	Ile	Ser	Arg	Asp	Glu	Ala	Val	His	Thr	Thr	Ala
			210					215				220				
	Ser	Cys	Tyr	Ile	Tyr	Asn	Asn	Tyr	Leu	Gly	Gly	His	Ala	Lys	Pro	Glu
	225					230					235				240	
	Ala	Ala	Arg	Val	Tyr	Arg	Leu	Phe	Arg	Glu	Ala	Val	Asp	Ile	Glu	Ile
30					245					250					255	
	Gly	Phe	Ile	Arg	Ser	Gln	Ala	Pro	Thr	Asp	Ser	Ser	Ile	Leu	Ser	Pro
				260					265					270		
	Gly	Ala	Ala	Ile	Glu	Asn	Tyr	Val	Arg	Phe	Ser	Ala	Asp	Arg	Leu	Leu
			275					280					285			
35	Gly	Leu	Ile	His	Met	Gln	Pro	Lys	Ala	Pro	Ala	Pro	Asp	Ala	Ser	Phe
		290						295				300				
	Pro	Leu	Ser	Leu	Met	Ser	Thr	Asp	Lys	His	Thr	Asn	Phe	Phe	Glu	Cys
	305					310					315				320	
	Arg	Ser	Thr	Ser	Tyr	Ala	Gly	Ala	Val	Val	Asn	Asp	Leu			
40					325						330					

(2) INFORMATION FOR SEQ ID NO:256:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 357 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

5

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

10

```

Met Arg Arg Arg Gly His Ala Phe Ala Pro Gly Asp Arg Gly Thr Arg
 1             5             10             15
Ala Ala Gly Pro Gly Pro Ala Ala Pro Trp Gly Ala Pro Ser Lys Pro
      20             25             30
15 Ala Leu Arg Leu Ala His Leu Phe Cys Ile Arg Val Leu Arg Ala Leu
      35             40             45
Gly Tyr Ala Tyr Ile Asn Ser Gly Gln Leu Glu Ala Asp Asp Ala Cys
      50             55             60
Ala Asn Leu Tyr His Thr Asn Thr Val Ala Tyr Val His Thr Thr Asp
20 65             70             75             80
Thr Asp Leu Leu Leu Met Gly Cys Asp Ile Val Leu Asp Ile Ser Thr
      85             90             95
Gly Tyr Ile Pro Thr Ile His Cys Arg Asp Leu Leu Gln Tyr Phe Lys
      100            105            110
25 Met Ser Tyr Pro Gln Phe Leu Ala Leu Phe Val Arg Cys His Thr Asp
      115            120            125
Leu His Pro Asn Asn Thr Tyr Ala Ser Val Glu Asp Val Leu Arg Glu
      130            135            140
Cys His Trp Thr Ala Pro Ser Arg Ser Gln Ala Arg Arg Ala Ala Arg
30 145            150            155            160
Arg Glu Arg Ala Asn Ser Arg Ser Leu Glu Ser Met Pro Thr Leu Thr
      165            170            175
Ala Ala Pro Val Gly Leu Glu Thr Arg Ile Ser Trp Thr Glu Ile Leu
      180            185            190
35 Ala Gln Gln Ile Ala Gly Glu Asp Asp Tyr Glu Glu Asp Pro Pro Leu
      195            200            205
Gln Pro Pro Asp Val Ala Gly Gly Pro Arg Asp Gly Ala Arg Ser Ser
      210            215            220
Ser Ser Glu Ile Leu Thr Pro Pro Glu Leu Val Gln Val Pro Asn Ala
40 225            230            235            240
Gln Arg Val Ala Glu His Arg Gly Tyr Val Ala Gly Arg Arg Arg His
      245            250            255
Val Ile His Asp Ala Pro Glu Ala Leu Asp Trp Leu Pro Asp Pro Met
645

```

260 265 270  
 Thr Ile Ala Glu Leu Val Glu His Arg Tyr Val Lys Tyr Val Ile Ser  
 275 280 285  
 Leu Ile Ser Pro Lys Glu Arg Gly Pro Trp Thr Leu Leu Lys Arg Leu  
 5 290 295 300  
 Pro Ile Tyr Gln Asp Leu Arg Asp Glu Asp Leu Ala Arg Ser Ile Val  
 305 310 315 320  
 Thr Arg His Ile Thr Ala Pro Asp Ile Ala Asp Arg Phe Leu Ala Gln  
 325 330 335  
 10 Leu Trp Ala His Ala Pro Pro Pro Ala Phe Tyr Lys Asp Val Leu Ala  
 340 345 350  
 Lys Phe Trp Asp Glu  
 355

15 (2) INFORMATION FOR SEQ ID NO:257:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 466 amino acids

(B) TYPE: amino acid

20 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

Met Ala His Leu Pro Gly Gly Ala Ala Ala Pro Leu Ser Glu Asp  
 1 5 10 15  
 Ala Ile Pro Ser Pro Arg Glu Arg Thr Glu Asp Trp Pro Pro Cys Gln  
 20 25 30  
 Ile Val Leu Gln Gly Ala Glu Leu Asn Gly Ile Leu Gln Ala Phe Ala  
 35 40 45  
 Pro Leu Arg Thr Ser Leu Leu Asp Ser Leu Leu Val Val Gly Asp Arg  
 50 55 60  
 35 Gly Ile Leu Val His Asn Ala Ile Phe Gly Glu Gln Val Phe Leu Pro  
 65 70 75 80  
 Leu Asp His Ser Gln Phe Ser Arg Tyr Arg Trp Gly Gly Pro Thr Ala  
 85 90 95  
 Ala Phe Leu Ser Leu Val Asp Gln Lys Arg Ser Leu Leu Ser Val Phe  
 100 105 110  
 40 Arg Ala Asn Gln Tyr Pro Asp Leu Arg Arg Val Glu Leu Thr Val Thr  
 115 120 125  
 Gly Gln Ala Pro Phe Arg Thr Leu Val Gln Arg Ile Trp Thr Thr Ala  
 646

	130		135		140	
	Ser Asp Gly Glu Ala Val	Glu Leu Ala Ser Glu Thr Leu Met Lys Arg				
	145	150	155	160		
	Glu Leu Thr Ser Phe Ala Val	Leu Leu Pro Gln Gly Asp Pro Asp Val				
5	165	170	175			
	Gln Leu Arg Leu Thr Lys Pro Gln Leu Thr Lys Val Val Asn Ala Val					
	180	185	190			
	Gly Asp Glu Thr Ala Lys Pro Thr Thr Phe Glu Leu Gly Pro Asn Gly					
	195	200	205			
10	Lys Phe Ser Val Phe Asn Ala Arg Thr Cys Val Thr Phe Ala Ala Arg					
	210	215	220			
	Glu Glu Gly Ala Ser Ser Ser Thr Ser Ala Gln Val Gln Ile Leu Thr					
	225	230	235	240		
	Ser Ala Leu Lys Lys Ala Gly Gln Ala Ala Ala Asn Ala Lys Thr Val					
15	245	250	255			
	Tyr Gly Glu Asn Thr Thr Phe Ser Val Val Val Asp Asp Cys Ser Met					
	260	265	270			
	Arg Ala Val Leu Arg Arg Leu Gln Val Gly Gly Gly Thr Leu Lys Phe					
	275	280	285			
20	Phe Leu Thr Ala Asp Val Pro Ser Val Cys Val Thr Ala Thr Gly Pro					
	290	295	300			
	Asn Ala Val Ser Ala Val Phe Leu Leu Lys Pro Gln Arg Val Cys Leu					
	305	310	315	320		
	Asn Trp Leu Gly Arg Thr Pro Gly Ser Ser Thr Gly Ser Leu Ala Ser					
25	325	330	335			
	Gln Asp Ser Arg Ala Gly Pro Thr Asp Ser Gln Asp Phe Ser Ser Glu					
	340	345	350			
	Pro Asp Ala Gly Asp Arg Gly Ala Pro Glu Glu Glu Gly Leu Glu Gly					
	355	360	365			
30	Gln Ala Arg Val Pro Pro Ala Phe Pro Glu Pro Pro Gly Thr Lys Arg					
	370	375	380			
	Arg His Ala Gly Ala Glu Val Val Pro Ala Asp Asp Ala Thr Lys Arg					
	385	390	395	400		
	Pro Lys Thr Gly Val Pro Ala Ala Pro Thr Arg Ala Glu Ser Pro Pro					
35	405	410	415			
	Leu Ser Ala Arg Tyr Gly Pro Glu Ala Ala Glu Gly Gly Gly Asp Gly					
	420	425	430			
	Gly Arg Tyr Ala Cys Tyr Phe Arg Asp Leu Gln Thr Gly Asp Asp Ser					
	435	440	445			
40	Pro Leu Ser Ala Phe Arg Gly Pro Gln Arg Pro Pro Tyr Gly Phe Gly					
	450	455	460			
	Leu Pro					
	465					

## (2) INFORMATION FOR SEQ ID NO:258:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 170 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 10 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

```

Met Ala Phe Arg Ala Ser Gly Pro Ala Tyr Gln Pro Leu Ala Pro Ala
15 1           5           10           15
Asp Ala Arg Ala Arg Val Pro Ala Val Ala Trp Ile Gly Val Gly Ala
           20           25           30
Ile Val Gly Ala Phe Ala Leu Val Ala Ala Leu Val Leu Val Pro Pro
           35           40           45
20 Arg Ser Ser Trp Gly Leu Ser Pro Cys Asp Ser Gly Trp Gln Glu Phe
           50           55           60
Asn Ala Gly Cys Val Ala Trp Asp Pro Thr Pro Val Glu His Glu Gln
65           70           75           80
Ala Val Gly Gly Cys Ser Ala Pro Ala Thr Leu Ile Pro Arg Ala Ala
25           85           90           95
Ala Lys His Leu Ala Ala Leu Thr Arg Val Gln Ala Glu Arg Ser Ser
           100          105          110
Gly Tyr Trp Trp Val Asn Gly Asp Gly Ile Arg Thr Cys Leu Arg Leu
           115          120          125
30 Val Asp Ser Val Ser Gly Ile Asp Glu Phe Cys Glu Glu Leu Ala Ile
           130          135          140
Arg Ile Cys Tyr Tyr Pro Arg Ser Pro Gly Gly Phe Val Arg Phe Val
145          150          155          160
Thr Ser Ile Arg Asn Ala Leu Gly Leu Pro
35           165           170

```

## (2) INFORMATION FOR SEQ ID NO:259:

## (i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 713 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

5  
Met Gln Arg Arg Arg Ala Ser Ser Leu Arg Leu Ala Arg Cys Leu Thr  
1 5 10 15  
Pro Ala Asn Leu Ile Arg Gly Ala Asn Ala Gly Val Pro Glu Arg Arg  
20 25 30  
10 Ile Phe Ala Gly Cys Leu Leu Pro Thr Pro Glu Gly Leu Leu Ser Ala  
35 40 45  
Ala Val Gly Val Leu Arg Gln Arg Ala Asp Asp Leu Gln Pro Ala Phe  
50 55 60  
Leu Thr Gly Ala Asp Arg Ser Val Arg Leu Ala Ala Arg His His Asn  
15 65 70 75 80  
Thr Val Pro Glu Ser Leu Ile Val Asp Gly Leu Ala Ser Asp Pro His  
85 90 95  
Tyr Asp Tyr Ile Arg His Tyr Ala Ser Ala Ala Lys Gln Ala Leu Gly  
100 105 110  
20 Glu Val Glu Leu Ser Gly Gly Gln Leu Ser Arg Ala Ile Leu Ala Gln  
115 120 125  
Tyr Trp Lys Tyr Leu Gln Thr Val Val Pro Ser Gly Leu Asp Ile Pro  
130 135 140  
Asp Asp Pro Ala Gly Asp Cys Asp Pro Ser Leu His Val Leu Leu Arg  
25 145 150 155 160  
Pro Thr Leu Leu Pro Lys Leu Leu Val Arg Ala Pro Phe Lys Ser Gly  
165 170 175  
Ala Ala Ala Ala Lys Tyr Ala Ala Ala Val Ala Gly Leu Arg Asp Ala  
180 185 190  
30 Ala His Arg Leu Gln Gln Tyr Met Phe Phe Met Arg Pro Ala Asp Pro  
195 200 205  
Ser Arg Pro Ser Thr Asp Thr Ala Leu Arg Leu Ser Glu Phe Leu Ala  
210 215 220  
Tyr Val Ser Val Leu Tyr His Trp Ala Ser Trp Met Leu Trp Thr Ala  
35 225 230 235 240  
Asp Lys Tyr Val Cys Arg Arg Leu Gly Pro Ala Asp Arg Arg Phe Val  
245 250 255  
Ser Gly Ser Leu Glu Ala Pro Ala Glu Thr Phe Ala Arg His Leu Asp  
260 265 270  
40 Arg Gly Pro Ser Gly Thr Thr Gly Ser Met Gln Cys Met Ala Leu Arg  
275 280 285  
Ala Ala Val Ser Asp Val Leu Gly His Leu Thr Arg Leu Ala His Leu  
290 295 300



	Trp	Glu	Thr	Gly	Lys	Arg	Ser	Gly	Gly	Thr	Tyr	Gly	Ile	Val	Asp	Ala
	305					310						315				320
	Ile	Val	Ser	Thr	Val	Glu	Val	Leu	Ser	Ile	Val	His	His	His	Ala	Gln
					325					330						335
5	Tyr	Ile	Ile	Asn	Ala	Thr	Leu	Thr	Gly	Tyr	Val	Val	Trp	Ala	Ser	Asp
					340				345					350		
	Ser	Leu	Asn	Asn	Glu	Tyr	Leu	Arg	Ala	Ala	Val	Asp	Ser	Gln	Glu	Arg
					355			360						365		
	Phe	Cys	Arg	Thr	Ala	Ala	Pro	Leu	Phe	Pro	Thr	Met	Thr	Ala	Pro	Ser
10		370					375					380				
	Trp	Ala	Arg	Met	Glu	Leu	Ser	Ile	Lys	Ser	Trp	Phe	Gly	Ala	Ala	Pro
	385					390					395					400
	Asp	Leu	Leu	Arg	Ser	Gly	Thr	Pro	Ser	Pro	His	Tyr	Glu	Ser	Ile	Leu
					405					410					415	
15	Arg	Leu	Ala	Ala	Ser	Gly	Pro	Pro	Gly	Gly	Arg	Gly	Ala	Val	Gly	Gly
					420				425					430		
	Ser	Cys	Arg	Asp	Lys	Ile	Gln	Arg	Thr	Arg	Arg	Asp	Asn	Ala	Pro	Pro
		435					440					445				
	Pro	Leu	Pro	Arg	Ala	Arg	Pro	His	Ser	Thr	Pro	Ala	Ala	Pro	Arg	Arg
20		450					455					460				
	Phe	Arg	Arg	His	Arg	Glu	Asp	Leu	Pro	Glu	Pro	Pro	His	Val	Asp	Ala
	465					470				475					480	
	Ala	Asp	Arg	Gly	Pro	Glu	Pro	Cys	Ala	Gly	Arg	Pro	Ala	Thr	Tyr	Tyr
					485					490					495	
25	Thr	His	Met	Ala	Gly	Ala	Pro	Pro	Arg	Leu	Pro	Pro	Arg	Asn	Pro	Ala
					500				505					510		
	Pro	Pro	Glu	Gln	Arg	Pro	Ala	Ala	Ala	Ala	Arg	Pro	Leu	Ala	Ala	Gln
			515					520					525			
	Arg	Glu	Ala	Ala	Gly	Val	Tyr	Asp	Ala	Val	Arg	Thr	Trp	Gly	Pro	Asp
30		530					535					540				
	Ala	Glu	Ala	Glu	Pro	Asp	Gln	Met	Glu	Asn	Thr	Tyr	Leu	Leu	Pro	Asp
	545					550				555					560	
	Asp	Asp	Ala	Ala	Met	Pro	Ala	Gly	Val	Gly	Leu	Gly	Ala	Thr	Pro	Ala
					565					570					575	
35	Ala	Asp	Thr	Thr	Ala	Ala	Ala	Trp	Pro	Ala	Glu	Ser	His	Ala	Pro	Arg
					580				585					590		
	Ala	Pro	Ser	Glu	Asp	Ala	Asp	Ser	Ile	Tyr	Glu	Ser	Val	Ser	Glu	Asp
					595			600					605			
	Gly	Gly	Arg	Val	Tyr	Glu	Glu	Ile	Pro	Trp	Val	Arg	Val	Tyr	Glu	Asn
40		610					615					620				
	Ile	Cys	Leu	Arg	Arg	Gln	Asp	Ala	Gly	Gly	Ala	Ala	Pro	Pro	Gly	Asp

645                      650                      655  
 Gly Gly Ser Ala Leu Phe Ser Pro Pro Gly Ala Thr Arg Ala Pro Asp  
                     660                      665                      670  
 Pro Gly Leu Ser Leu Ser Pro Met Pro Ala Arg Pro Arg Thr Asn Ala  
 5                      675                      680                      685  
 Asn Asp Gly Pro Thr Asn Val Ala Ala Leu Ser Ala Leu Leu Thr Lys  
                     690                      695                      700  
 Leu Lys Arg Gly Arg His Gln Ser His  
 705                      710

10

## (2) INFORMATION FOR SEQ ID NO:260:

## (i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 352 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

Val Gly Ala Ala Ala Val Pro Leu Leu Ser Ala Gly Gly Ala Ala Pro  
 1                      5                      10                      15  
 25 Pro His Pro Gly Pro Asp Ala Ala Val Phe Arg Ser Ser Leu Gly Ser  
                     20                      25                      30  
 Leu Leu Tyr Trp Pro Gly Val Arg Ala Leu Leu Gly Arg Asp Cys Arg  
                     35                      40                      45  
 Val Ala Ala Arg Tyr Ala Gly Arg Met Thr Tyr Ile Ala Thr Gly Ala  
 30                      50                      55                      60  
 Leu Leu Ala Arg Phe Asn Pro Gly Ala Val Lys Cys Val Leu Pro Arg  
 65                      70                      75                      80  
 Glu Ala Ala Phe Ala Gly Arg Val Leu Asp Val Leu Ala Val Leu Ala  
                     85                      90                      95  
 35 Glu Gln Thr Val Gln Trp Leu Ser Val Val Val Gly Ala Arg Leu His  
                     100                      105                      110  
 Pro His Ser Ala His Pro Ala Phe Val Asp Val Glu Gln Glu Ala Leu  
                     115                      120                      125  
 Phe Arg Ala Leu Pro Leu Gly Ser Pro Gly Val Val Ala Ala Glu His  
 40                      130                      135                      140  
 Glu Ala Leu Gly Asp Thr Ala Ala Arg Arg Leu Leu Ala Thr Ser Gln  
 145                      150                      155                      160  
 Ala Val Leu Gly Ala Ala Val Tyr Ala Leu His Thr Ala Thr Val Thr

165 170 175  
 Leu Lys Tyr Ala Cys Gly Asp Ala Arg Arg Arg Arg Asp His Ala Ala  
 180 185 190  
 Ala Ala Arg Ala Val Leu Ala Thr Gly Leu Ile Leu Gln Arg Leu Leu  
 5 195 200 205  
 Gly Leu Ala Asp Thr Val Val Ala Cys Val Ala Ala Phe Asp Gly Gly  
 210 215 220  
 Ser Thr Ala Pro Glu Val Gly Thr Tyr Thr Pro Leu Arg Tyr Ala Cys  
 225 230 235 240  
 10 Val Leu Arg Ala Thr Gln Pro Leu Tyr Ala Arg Thr Thr Pro Ala Lys  
 245 250 255  
 Phe Trp Ala Asp Val Arg Ala Ala Ala Glu His Val Asp Leu Arg Pro  
 260 265 270  
 Ala Ser Ser Ala Pro Arg Ala Pro Val Ser Gly Thr Ala Asp Pro Ala  
 15 275 280 285  
 Phe Leu Leu Glu Asp Leu Ala Ala Phe Pro Pro Ala Pro Leu Asn Ser  
 290 295 300  
 Glu Ser Val Leu Gly Pro Arg Val Arg Val Val Asp Ile Met Ala Gln  
 305 310 315 320  
 20 Phe Arg Lys Leu Leu Met Gly Asp Glu Glu Thr Ala Ala Leu Arg Ala  
 325 330 335  
 His Val Ser Gly Arg Arg Ala Thr Gly Leu Gly Gly Pro Pro Arg Pro  
 340 345 350

25 (2) INFORMATION FOR SEQ ID NO:261:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 457 amino acids  
 (B) TYPE: amino acid  
 30 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

Met Ser Val Arg Gly His Ala Val Arg Arg Arg Arg Ala Ser Thr Arg  
 1 5 10 15  
 Ser His Ala Pro Ser Ala His Arg Ala Asp Ser Pro Val Glu Asp Glu  
 40 20 25 30  
 Pro Glu Gly Gly Gly Val Gly Leu Met Gly Tyr Leu Arg Ala Val Phe  
 35 40 45  
 Asn Val Asp Asp Asp Ser Glu Val Glu Ala Ala Gly Glu Met Ala Ser

	50		55		60											
	Glu	Glu	Pro	Pro	Pro	Arg	Arg	Arg	Arg	Glu	Arg	His	Pro	Gly	Ser	Arg
	65				70					75					80	
	Arg	Ala	Ser	Glu	Ala	Arg	Ala	Ala	Ala	Pro	Pro	Arg	Arg	Ala	Ser	Phe
5				85						90					95	
	Pro	Arg	Pro	Arg	Ser	Val	Thr	Ala	Arg	Ser	Gln	Ser	Val	Arg	Gly	Arg
				100						105					110	
	Arg	Asp	Ser	Ala	Ile	Thr	Arg	Ala	Pro	Arg	Gly	Gly	Tyr	Leu	Gly	Pro
				115						120					125	
10	Met	Asp	Pro	Arg	Asp	Val	Leu	Gly	Arg	Val	Gly	Gly	Ser	Arg	Val	Val
	130									135					140	
	Pro	Ser	Pro	Leu	Phe	Leu	Asp	Glu	Leu	Ser	Tyr	Glu	Glu	Asp	Asp	Tyr
	145					150					155				160	
	Pro	Ala	Ala	Val	Ala	His	Asp	Asp	Gly	Ala	Gly	Ala	Arg	Pro	Pro	Ala
15				165						170					175	
	Thr	Val	Glu	Ile	Leu	Ala	Gly	Arg	Val	Ser	Gly	Pro	Glu	Leu	Gln	Ala
				180						185					190	
	Ala	Phe	Pro	Leu	Asp	Arg	Leu	Thr	Pro	Arg	Val	Ala	Ala	Trp	Asp	Glu
				195						200					205	
20	Ser	Val	Arg	Ser	Ala	Leu	Gly	His	Pro	Ala	Gly	Phe	Tyr	Pro	Cys	Pro
	210									215					220	
	Asp	Ser	Ala	Phe	Gly	Leu	Ser	Arg	Val	Gly	Val	Met	His	Phe	Asp	Ala
	225					230					235				240	
	Asp	Pro	Lys	Val	Phe	Phe	Arg	Gln	Thr	Leu	Gln	Gln	Gly	Glu	Ala	Trp
25				245						250					255	
	Tyr	Val	Thr	Gly	Asp	Ala	Ile	Leu	Asp	Leu	Thr	Asp	Arg	Arg	Ala	Lys
				260						265					270	
	Thr	Ser	Pro	Ser	Arg	Ala	Met	Gly	Phe	Leu	Val	Asp	Ala	Ile	Val	Arg
				275						280					285	
30	Val	Ala	Ile	Asn	Gly	Trp	Val	Cys	Gly	Thr	Arg	Leu	His	Thr	Glu	Gly
	290									295					300	
	Ala	Arg	Leu	Gly	Ala	Arg	Arg	Gln	Gly	Gly	Arg	Ala	Pro	Thr	Ala	Val
	305					310					315				320	
	Arg	Glu	Pro	His	Gly	Val	Ala	Arg	Gly	Arg	Arg	Arg	Ala	Ala	Ala	Gln
35				325						330					335	
	Arg	Gly	Arg	Gly	Arg	Ala	Pro	Pro	Pro	Arg	Pro	Arg	Arg	Arg	Gly	Leu
				340						345					350	
	Ser	Gln	Phe	Ala	Gly	Val	Pro	Ala	Val	Leu	Arg	Ala	Arg	Ala	Pro	Gly
				355						360					365	
40	Ala	Arg	Leu	Ser	Arg	Gly	Arg	Pro	Leu	Arg	Gly	Ala	His	Asp	Val	His
	370									375					380	
	Arg	His	Arg	Gly	Ser	Ala	Arg	Pro	Leu	Gln	Pro	Arg	Arg	Arg	Gln	Met
	385					390					395				400	

Arg Ala Pro Ala Gly Gly Arg Val Cys Gly Ala Arg Pro Gly Arg Ala  
                             405                            410                            415  
 Gly Gly Pro Gly Gly Ala Asp Gly Pro Val Gly Gly Arg Gly Gly Ala  
                             420                            425                            430  
 5 Pro Ala Pro Ala Leu Arg Pro Pro Arg Val Cys Gly Arg Gly Ala Gly  
                             435                            440                            445  
 Gly Ala Val Ser Arg Pro Ala Pro Gly  
                             450                            455

10 (2) INFORMATION FOR SEQ ID NO:262:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 298 amino acids  
 (B) TYPE: amino acid  
 15 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

Met Thr Ser Arg Arg Ser Val Lys Ser Cys Pro Arg Glu Ala Pro Arg  
   1                            5                            10                            15  
 Gly Thr His Glu Glu Leu Tyr Tyr Gly Pro Val Ser Pro Ala Asp Pro  
 25                            20                            25                            30  
 Glu Ser Pro Arg Asp Asp Phe Arg Arg Gly Ala Gly Pro Met Arg Ala  
                             35                            40                            45  
 Arg Pro Arg Gly Glu Val Arg Phe Leu His Tyr Asp Glu Ala Gly Tyr  
                             50                            55                            60  
 30 Ala Leu Tyr Arg Asp Ser Ser Ser Ser Glu Asp Asn Asp Glu Ser Arg  
                             65                            70                            75                            80  
 Asp Thr Ala Arg Pro Arg Arg Ser Ala Ser Val Ala Gly Ser His Gly  
                             85                            90                            95  
 Pro Gly Pro Ala Arg Ala Pro Pro Pro Pro Gly Gly Pro Val Gly Ala  
 35                            100                            105                            110  
 Gly Gly Arg Ser His Ala Pro Pro Ala Arg Thr Pro Lys Met Thr Arg  
                             115                            120                            125  
 Gly Ala Pro Lys Ala Pro Ala Thr Pro Ala Thr Asp Pro Arg Arg Arg  
                             130                            135                            140  
 40 Pro Ala Gln Ala Asp Ser Ala Val Leu Leu Asp Ala Pro Ala Pro Thr  
                             145                            150                            155                            160  
 Ala Ser Gly Arg Thr Lys Thr Pro Ala Gln Gly Leu Ala Lys Lys Leu  
                             165                            170                            175

His Phe Ser Thr Ala Pro Pro Ser Pro Thr Ala Pro Trp Thr Pro Arg  
 180 185 190  
 Val Ala Gly Phe Asn Lys Arg Val Phe Cys Ala Ala Val Gly Arg Leu  
 195 200 205  
 5 Ala Ala Thr His Ala Arg Leu Ala Ala Val Gln Leu Trp Asp Met Ser  
 210 215 220  
 Arg Pro His Thr Asp Glu Asp Leu Asn Glu Leu Leu Asp Leu Thr Thr  
 225 230 235 240  
 Ile Arg Val Thr Val Cys Glu Gly Lys Asn Leu Leu Gln Arg Ala Asn  
 10 245 250 255  
 Glu Leu Val Asn Pro Asp Ala Ala Gln Asp Val Asp Ala Thr Ala Ala  
 260 265 270  
 Arg Arg Pro Ala Gly Arg Ala Ala Ala Thr Ala Arg Ala Pro Ala Arg  
 275 280 285  
 15 Ser Ala Ser Arg Pro Arg Arg Pro Leu Glu  
 290 295

## (2) INFORMATION FOR SEQ ID NO:263:

## 20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 70 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

25

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

30 Val Val Leu Leu Phe Val Val Ala Gly Val Pro Gly Glu Pro Pro Asn  
 1 5 10 15  
 Ala Ala Gly Arg Val Ile Gly Asp Ala Gln Cys Arg Gly Asp Ser Ala  
 20 25 30  
 Gly Val Val Ser Val Pro Gly Val Leu Val Pro Phe Tyr Leu Gly Met  
 35 35 40 45  
 Thr Ser Met Gly Val Cys Met Ile Ala His Val Tyr Gln Ile Cys Gln  
 50 55 60  
 Arg Ala Ala Gly Ser Ala  
 65 70

40

## (2) INFORMATION FOR SEQ ID NO:264:

## (i) SEQUENCE CHARACTERISTICS:

655

(A) LENGTH: 363 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:264:

```

10 Met Ser Gln Trp Gly Pro Arg Ala Ile Leu Val Gln Thr Asp Ser Thr
    1           5           10           15
    Asn Arg Asn Ala Asp Gly Asp Trp Gln Ala Ala Val Ala Ile Arg Gly
        20           25           30
    Gly Gly Val Val Gln Leu Asn Met Val Asn Lys Arg Ala Val Asp Phe
15           35           40           45
    Thr Pro Ala Glu Cys Gly Asp Ser Glu Trp Ala Val Gly Arg Val Ser
    50           55           60
    Leu Gly Leu Arg Met Ala Met Pro Arg Asp Phe Cys Ala Ile Ile His
    65           70           75           80
20 Ala Pro Ala Val Ser Gly Pro Gly Pro His Val Met Leu Gly Leu Val
        85           90           95
    Asp Ser Gly Tyr Arg Gly Thr Val Leu Ala Val Val Val Ala Pro Asn
        100          105          110
    Gly Thr Arg Gly Phe Ala Pro Gly Ala Leu Arg Val Asp Val Thr Phe
25           115          120          125
    Leu Asp Ile Arg Ala Thr Pro Pro Thr Leu Thr Glu Pro Ser Ser Leu
    130          135          140
    His Arg Phe Pro Gln Leu Ala Pro Ser Pro Leu Ala Gly Leu Arg Glu
    145          150          155          160
30 Asp Pro Trp Leu Asp Gly Ala Thr Ala Gly Gly Ala Val Pro Ala Arg
        165          170          175
    Arg Arg Gly Gly Ser Leu Val Tyr Ala Gly Glu Leu Thr Gln Val Thr
        180          185          190
    Thr Glu His Gly Asp Cys Val His Glu Ala Pro Ala Phe Leu Pro Lys
35           195          200          205
    Arg Glu Glu Asp Ala Gly Phe Asp Ile Leu Ile His Arg Ala Val Thr
    210          215          220
    Val Pro Ala Asn Gly Ala Thr Val Ile Gln Pro Ser Leu Arg Val Leu
    225          230          235          240
40 Arg Ala Ala Asp Gly Pro Glu Ala Cys Tyr Val Leu Gly Arg Ser Ser
        245          250          255
    Leu Asn Arg Leu Leu Val Met Pro Thr Arg Trp Pro Ser Gly His Ala
        260          265          270

```

Cys Ala Phe Val Val Cys Asn Leu Thr Gly Val Pro Val Thr Leu Gln  
 275 280 285  
 Ala Gly Ser Lys Val Ala Gln Leu Leu Val Ala Gly Thr His Ala Leu  
 290 295 300  
 5 Pro Trp Ile Pro Pro Asp Asn Ile His Glu Asp Gly Ala Phe Arg Ala  
 305 310 315 320  
 Tyr Pro Arg Gly Val Pro Asp Ala Thr Ala Thr Pro Arg Asp Pro Pro  
 325 330 335  
 Ile Leu Val Phe Thr Asn Glu Phe Asp Ala Asp Ala Pro Pro Ser Lys  
 10 340 345 350  
 Arg Gly Ala Gly Gly Phe Gly Ser Thr Gly Ile  
 355 360

## (2) INFORMATION FOR SEQ ID NO:265:

15

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 236 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 20 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:265:

25

Met Ala Ser Leu Leu Gly Val Leu Cys Gly Trp Gly Trp Glu Glu Gln  
 1 5 10 15  
 Gln Tyr Glu Met Ile Arg Ala Ala Ala Pro Pro Ser Xaa Xaa Asp Pro  
 20 25 30  
 30 Arg Leu Gln Glu Ala Val Val Asn Ala Leu Leu Pro Ala Pro Ile Thr  
 35 40 45  
 Leu Asp Asp Ala Leu Glu Ser Leu Asp Asp Thr Arg Arg Leu Val Lys  
 50 55 60  
 Ala Arg Ala Arg Thr Tyr His Ala Cys Met Val Asn Leu Glu Arg Leu  
 35 65 70 75 80  
 Ala Arg His His Pro Gly Leu Glu Gly Ser Thr Ile Asp Gly Ala Val  
 85 90 95  
 Ala Ala His Arg Asp Lys Met Arg Arg Leu Ala Asp Thr Cys Met Ala  
 100 105 110  
 40 Thr Ile Leu Gln Met Tyr Met Ser Val Gly Ala Ala Asp Lys Ser Ala  
 115 120 125  
 Asp Val Leu Val Ser Gln Ala Ile Arg Ser Met Ala Glu Ser Asp Val  
 130 135 140



Val Met Glu Asp Val Ala Ile Ala Glu Arg Ala Leu Gly Leu Ser Thr  
 145                      150                      155                      160  
 Ser Ala Gly Gly Thr Arg Thr Ala Gly Leu Gly Ala Thr Glu Ala Pro  
                          165                      170                      175  
 5    Pro Gly Pro Thr Arg Ala Gln Ala Pro Glu Val Ala Ser Val Pro Val  
                          180                      185                      190  
 Thr His Ala Gly Asp Arg Ser Pro Val Arg Pro Gly Pro Val Pro Pro  
                          195                      200                      205  
 Ala Asp Pro Thr Pro Asp Pro Arg His Arg Thr Ser Ala Pro Lys Arg  
 10                      210                      215                      220  
 Gln Ala Ser Ser Thr Glu Ala Pro Leu Leu Leu Ala  
 225                      230                      235

## (2) INFORMATION FOR SEQ ID NO:266:

15

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 453 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

20

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:

25

Met Tyr Val Asn Arg Asn Glu Ile Phe Asn Ala Ala Val Thr Asn Ile  
 1                      5                      10                      15  
 Ile Leu Asp Leu Asp Ile Ala Leu Lys Glu Pro Val Pro Phe Pro Arg  
                          20                      25                      30  
 30    Leu His Glu Ala Leu Gly His Phe Arg Arg Gly Ala Ala Val Gln Leu  
                          35                      40                      45  
 Leu Phe Pro Ala Ala Arg Val Asp Pro Asp Ala Tyr Pro Cys Tyr Phe  
                          50                      55                      60  
 Phe Lys Ser Ala Cys Arg Pro Arg Ala Pro Pro Val Cys Ala Gly Asp  
 35                      65                      70                      75                      80  
 Gly Pro Ser Ala Gly Gly Asp Asp Gly Asp Gly Asp Trp Phe Pro Asp  
                          85                      90                      95  
 Ala Gly Gly Asp Asp Gly Asp Glu Glu Trp Glu Glu Asp Thr Asp Pro  
                          100                      105                      110  
 40    Met Asp Thr Thr His Gly Pro Leu Pro Asp Asp Glu Ala Ala Tyr Leu  
                          115                      120                      125  
 Asp Leu Leu His Glu Gln Ile Pro Ala Ala Thr Pro Ser Glu Pro Asp  
                          130                      135                      140

Ser Val Val Cys Ser Cys Ala Asp Lys Ile Gly Leu Arg Val Cys Leu  
 145 150 155 160  
 Pro Val Pro Ala Pro Tyr Val Val His Gly Ser Leu Thr Met Arg Gly  
 165 170 175  
 5 Val Ala Arg Val Ile Gln Gln Ala Val Leu Leu Asp Arg Asp Phe Val  
 180 185 190  
 Glu Ala Val Gly Ser His Val Lys Asn Phe Leu Leu Ile Asp Thr Gly  
 195 200 205  
 Val Tyr Ala His Gly His Ser Leu Arg Leu Pro Tyr Phe Ala Lys Ile  
 10 210 215 220  
 Gly Pro Asp Gly Ser Ala Cys Gly Arg Leu Leu Pro Val Phe Val Ile  
 225 230 235 240  
 Pro Pro Ala Cys Glu Asp Val Pro Ala Phe Val Ala Ala His Ala Asp  
 245 250 255  
 15 Pro Arg Arg Phe His Phe His Ala Pro Pro Met Phe Ser Ala Ala Pro  
 260 265 270  
 Arg Glu Ile Arg Val Leu His Ser Leu Gly Gly Asp Tyr Val Ser Phe  
 275 280 285  
 Phe Glu Lys Lys Ala Ser Arg Asn Ala Leu Glu His Phe Gly Arg Arg  
 20 290 295 300  
 Glu Thr Leu Thr Glu Val Leu Gly Arg Tyr Asp Val Arg Pro Asp Ala  
 305 310 315 320  
 Gly Glu Thr Val Glu Gly Phe Ala Ser Glu Leu Leu Gly Arg Ile Val  
 325 330 335  
 25 Ala Cys Ile Glu Ala His Phe Pro Glu His Ala Arg Glu Tyr Gln Ala  
 340 345 350  
 Val Ser Val Arg Arg Ala Val Ile Lys Asp Asp Trp Val Leu Leu Gln  
 355 360 365  
 Leu Ile Pro Gly Arg Gly Ala Leu Asn Gln Ser Leu Ser Cys Leu Arg  
 30 370 375 380  
 Phe Lys His Gly Arg Ala Ser Arg Ala Thr Ala Arg Thr Phe Leu Ala  
 385 390 395 400  
 Leu Ser Val Gly Thr Asn Asn Arg Leu Cys Ala Ser Leu Cys Gln Gln  
 405 410 415  
 35 Cys Phe Ala Thr Lys Cys Asp Asn Asn Arg Leu His Thr Leu Phe Thr  
 420 425 430  
 Val Asp Ala Gly Thr Pro Cys Ser Arg Ser Ala Pro Ser Ser Thr Ser  
 435 440 445  
 Arg Pro Ser Ser Ser  
 40 450

(2) INFORMATION FOR SEQ ID NO:267:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 332 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:

10

Met Leu Ala Val Arg Ser Leu Gln His Leu Thr Thr Val Ile Phe Ile  
 1 5 10 15  
 Thr Ala Tyr Gly Leu Val Leu Ala Trp Tyr Ile Val Phe Gly Asp Leu  
 20 25 30  
 15 His Arg Cys Ile Tyr Ala Val Arg Pro Ala Gly Ala His Asn Asp Thr  
 35 40 45  
 Ala Leu Val Trp Met Lys Ile Asn Gln Thr Leu Leu Phe Leu Gly Pro  
 50 55 60  
 Pro Thr Ala Pro Pro Gly Gly Ala Trp Thr Pro His Ala His Val Cys  
 20 65 70 75 80  
 Tyr Ala Asn Ile Ile Glu Gly Arg Ala Val Ser Leu Pro Ala Ile Pro  
 85 90 95  
 Gly Ala Met Ser Arg Arg Val Met Asn Val His Glu Ala Val Asn Cys  
 100 105 110  
 25 Leu Glu Ala Leu Trp Asp Thr Gln Met Arg Leu Val Val Val Gly Trp  
 115 120 125  
 Phe Leu Tyr Leu Ala Phe Val His Gln Arg Arg Cys Met Phe Gly Val  
 130 135 140  
 Val Ser Pro Ala His Ser Met Val Ala Pro Ala Thr Tyr Leu Leu Asn  
 30 145 150 155 160  
 Tyr Ala Gly Arg Ile Val Ser Ser Val Phe Leu Gln Tyr Pro Tyr Thr  
 165 170 175  
 Lys Ile Thr Arg Leu Leu Cys Glu Leu Ser Val Gln Arg Gln Thr Leu  
 180 185 190  
 35 Val Gln Leu Phe Glu Ala Asp Pro Val Thr Phe Leu Tyr His Arg Pro  
 195 200 205  
 Ala Val Gly Val Ile Val Gly Cys Glu Leu Leu Leu Arg Phe Val Gly  
 210 215 220  
 Leu Ile Val Gly Thr Ala Leu Ile Ser Arg Gly Ala Cys Ala Ile Thr  
 40 225 230 235 240  
 Tyr Pro Leu Phe Leu Thr Ile Thr Thr Trp Cys Phe Val Ser Ile Ile  
 245 250 255  
 Ala Leu Thr Glu Leu Tyr Phe Ile Leu Arg Arg Asp Ser Ala Pro Lys

260 265 270  
 Asn Ala Glu Pro Ala Ala Pro Arg Gly Arg Ser Lys Gly Trp Ser Gly  
 275 280 285  
 Val Cys Gly Arg Cys Cys Ser Ile Ile Leu Ser Gly Ile Ala Val Arg  
 5 290 295 300  
 Leu Cys Tyr Ile Ala Val Val Ala Gly Val Val Leu Met Ala Leu Arg  
 305 310 315 320  
 Tyr Glu Gln Glu Ile Gln Arg Arg Leu Phe Asp Leu  
 325 330

10

## (2) INFORMATION FOR SEQ ID NO:268:

## (i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 117 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:

Met Ile Gly Ala His Pro Gly Val Gly Gly Asp Leu Pro Ser Gly Leu  
 1 5 10 15  
 25 Pro Thr Tyr Ala Glu Ala Thr Ser Asp Arg Pro Pro Thr Tyr Ala Met  
 20 25 30  
 Val Met Ala Ala Cys Pro Thr Glu Pro Pro Gly Gly Ser Val Gly Pro  
 35 40 45  
 Ala Asp Gln Pro Arg Val Gln Ser Ser Arg Thr Trp Arg Pro Pro Leu  
 30 50 55 60  
 Val Asn Ser Arg Glu Leu Tyr Arg Ala Gln Arg Ala Ala Arg Cys Ala  
 65 70 75 80  
 Ser Ser Ser Asp Thr Pro Gln Ala Pro Gly Trp Cys Gly Gly Thr Cys  
 85 90 95  
 35 Arg His Ala Val Phe Gly Val Val Ala Val Val Val Val Ile Ile Leu  
 100 105 110  
 Ala Phe Leu Trp Arg  
 115

40

## (2) INFORMATION FOR SEQ ID NO:269:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 194 amino acids

(B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:

10	Met	His	Leu	Phe	Cys	Gln	Cys	Pro	Leu	Thr	Asp	Gly	Gln	Asp	Leu	Tyr
	1				5					10					15	
	Leu	Cys	Pro	Val	Tyr	Pro	Arg	Met	His	Gln	Glu	His	Leu	Val	Cys	Pro
				20					25					30		
	Leu	His	Arg	Leu	Asp	Asp	Ala	Arg	Arg	Arg	Gly	Arg	Thr	Ser	Ala	Ala
				35				40					45			
15	Trp	Asp	Glu	Gly	Leu	Val	Arg	Ala	Leu	Thr	His	Ser	Gly	Gly	Leu	Met
		50					55					60				
	Gly	Cys	Gly	Gly	Arg	Ser	Leu	Thr	Leu	Ser	Glu	Thr	Tyr	Trp	Gly	His
	65				70					75				80		
	Pro	Leu	Tyr	Glu	Lys	Leu	Val	Pro	Trp	Asp	His	Pro	Arg	Asp	Leu	Lys
20					85				90					95		
	Val	Pro	Glu	Ala	Ser	Ala	Val	Gly	Thr	Arg	Ala	Leu	Val	Pro	Arg	Gly
				100					105					110		
	Arg	Gly	Arg	Pro	Leu	Arg	Gly	Arg	Pro	Val	Pro	Leu	Ile	Pro	Leu	Asp
				115				120					125			
25	Cys	Glu	Pro	Asn	Asp	Gly	Leu	Pro	Phe	Gly	Gly	Gly	Trp	Pro	Gly	Gly
		130					135					140				
	Arg	Leu	Arg	Gly	Ala	Pro	Val	Pro	Leu	His	Pro	Pro	Pro	Pro	Ser	Ala
	145					150					155				160	
	Pro	Pro	Leu	Ser	Phe	Thr	Pro	Thr	Leu	Thr	Pro	Pro	Cys	Leu	Cys	Arg
30					165				170					175		
	Gly	Leu	Ser	Leu	Cys	Val	Val	Val	Lys	Gln	Tyr	Leu	Lys	Asp	Arg	Asn
				180					185					190		
	Asn	Phe														

35

(2) INFORMATION FOR SEQ ID NO:270:

(i) SEQUENCE CHARACTERISTICS:

40 (A) LENGTH: 853 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:

```

5  Met Asn Val Ala Thr Cys Thr His Gln Thr His His Ala Ala Arg Ala
    1             5             10             15
    Pro Gly Ala Thr Ser Ala Pro Gly Ala Ala Ser Gly Asp Pro Leu Gly
          20             25             30
    Ala Arg Arg Pro Ile Gly Asp Asp Glu Cys Glu Gln Tyr Thr Ser Ser
10          35             40             45
    Val Ser Leu Ala Arg Met Leu Tyr Gly Gly Asp Leu Ala Glu Trp Val
    50             55             60
    Pro Arg Val His Pro Lys Thr Thr Ile Glu Arg Gln Gln His Gly Pro
    65             70             75             80
15  Val Thr Phe Pro Asp Ala Ser Ala Pro Thr Ala Arg Cys Val Thr Val
          85             90             95
    Val Arg Ala Pro Met Gly Ser Gly Lys Thr Thr Ala Leu Ile Arg Trp
          100            105            110
    Leu Gly Glu Ala Ile His Ser Pro Asp Thr Ser Val Leu Val Val Ser
20          115            120            125
    Cys Arg Arg Ser Phe Thr Gln Thr Leu Ala Thr Arg Phe Ala Glu Ser
    130            135            140
    Gly Leu Pro Asp Phe Val Thr Tyr Phe Ser Ser Thr Asn Tyr Ile Met
    145            150            155            160
25  Asn Asp Arg Pro Phe His Arg Leu Ile Val Gln Val Glu Ser Leu His
          165            170            175
    Arg Val Gly Pro Asn Leu Leu Asn Asn Tyr Asp Val Leu Val Leu Asp
          180            185            190
    Glu Val Met Ser Thr Leu Gly Gln Lys Pro Thr Met Gln Gln Leu Gly
30          195            200            205
    Arg Val Asp Ala Leu Met Leu Arg Leu Leu Arg Thr Cys Pro Arg Ile
    210            215            220
    Ile Ala Met Asp Ala Thr Ala Asn Ala Gln Leu Val Asp Phe Leu Cys
    225            230            235            240
35  Ser Leu Arg Gly Glu Lys Asn Val His Val Val Ile Gly Glu Tyr Ala
          245            250            255
    Met Pro Gly Phe Ser Ala Arg Arg Cys Leu Phe Leu Pro Arg Leu Gly
          260            265            270
    Pro Glu Val Leu Gln Ala Ala Leu Arg Pro Pro Gly Pro Ala Gly Gly
40          275            280            285
    Ala Pro Pro Pro Asp Ala Pro Pro Asp Ala Thr Phe Phe Gly Glu Leu
    290            295            300
    Glu Ala Arg Leu Ala Gly Gly Asp Asn Val Cys Ile Phe Ser Ser Thr

```

	305					310						315				320
	Val	Ser	Phe	Ala	Glu	Val	Val	Ala	Arg	Phe	Cys	Arg	Gln	Phe	Thr	Asp
					325					330						335
5	Arg	Val	Leu	Leu	Leu	His	Ser	Leu	Thr	Pro	Pro	Gly	Asp	Val	Thr	Thr
				340					345					350		
	Trp	Gly	Arg	Tyr	Arg	Val	Val	Ile	Tyr	Thr	Thr	Val	Val	Thr	Val	Gly
			355					360					365			
	Leu	Ser	Phe	Asp	Pro	Pro	His	Phe	Asp	Ser	Met	Phe	Ala	Tyr	Val	Lys
		370					375				380					
10	Pro	Met	Asn	Tyr	Gly	Pro	Asp	Met	Val	Ser	Val	Tyr	Gln	Ser	Leu	Gly
	385					390					395					400
	Arg	Val	Arg	Thr	Leu	Arg	Lys	Gly	Glu	Leu	Leu	Ile	Tyr	Met	Asp	Gly
				405					410						415	
	Ser	Gly	Ala	Arg	Ser	Glu	Pro	Val	Phe	Thr	Pro	Met	Leu	Leu	Asn	His
15			420						425				430			
	Val	Val	Ser	Ala	Ser	Gly	Gln	Trp	Pro	Ala	Gln	Phe	Ser	Gln	Val	Thr
			435					440					445			
	Asn	Leu	Leu	Cys	Arg	Arg	Phe	Lys	Gly	Arg	Cys	Asp	Ala	Ser	His	Ala
		450					455				460					
20	Asp	Ala	Ala	Gln	Arg	Ser	Arg	Ile	Tyr	Ser	Lys	Phe	Arg	Tyr	Lys	His
	465					470					475					480
	Tyr	Phe	Glu	Arg	Cys	Thr	Leu	Ala	Cys	Leu	Ala	Asp	Ser	Leu	Asn	Ile
				485					490						495	
	Leu	His	Met	Leu	Leu	Thr	Leu	Asn	Cys	Met	His	Val	Arg	Phe	Trp	Gly
25			500						505				510			
	His	Asp	Ala	Ala	Leu	Thr	Pro	Arg	Asn	Phe	Cys	Leu	Phe	Leu	Arg	Gly
		515					520					525				
	Ile	His	Phe	Asp	Ala	Leu	Arg	Ala	Gln	Arg	Asp	Leu	Arg	Glu	Leu	Arg
		530					535				540					
30	Cys	Gln	Asp	Pro	Asp	Thr	Ser	Leu	Ser	Ala	Gln	Ala	Ala	Glu	Thr	Glu
	545					550					555					560
	Glu	Val	Gly	Leu	Phe	Val	Glu	Lys	Tyr	Leu	Arg	Pro	Asp	Val	Ala	Pro
				565					570						575	
	Ala	Glu	Val	Val	Met	Arg	Gln	Ser	Leu	Val	Gly	Arg	Thr	Arg	Phe	Ile
35			580						585				590			
	Tyr	Leu	Val	Leu	Leu	Glu	Ala	Cys	Leu	Arg	Val	Pro	Met	Ala	Ala	His
		595						600				605				
	Ser	Ser	Ala	Ile	Phe	Arg	Arg	Leu	Tyr	Asp	His	Tyr	Ala	Thr	Gly	Val
		610					615				620					
40	Ile	Pro	Thr	Ile	Asn	Ala	Ala	Gly	Glu	Leu	Glu	Leu	Val	His	Pro	Thr
	625					630					635					640
	Leu	Asn	Val	Al												

Ala Ala Cys Leu Gln Trp Asp Ser Met Ala Gly Gly Ser Gly Arg Thr  
660 665 670  
Phe Ser Pro Glu Asp Val Leu Glu Leu Leu Asn Pro His Tyr Asp Arg  
675 680 685  
5 Tyr Met Gln Leu Val Phe Glu Leu Gly His Cys Asn Val Thr Asp Gly  
690 695 700  
Pro Leu Leu Ser Glu Asp Ala Val Lys Arg Val Ala Asp Ala Leu Ser  
705 710 715 720  
Gly Cys Pro Pro Arg Gly Ser Val Ser Glu Thr Glu His Ala Leu Ser  
10 725 730 735  
Leu Phe Lys Ile Ile Trp Gly Glu Leu Phe Gly Val Gln Leu Ala Lys  
740 745 750  
Ser Thr Gln Thr Phe Pro Gly Ala Gly Arg Val Lys Asn Leu Thr Lys  
755 760 765  
15 Arg Ala Ile Val Glu Leu Leu Asp Ala His Arg Ile Asp His Ser Ala  
770 775 780  
Cys Arg Thr Gln Leu Tyr Ala Leu Leu Met Ala His Lys Arg Glu Phe  
785 790 795 800  
Ala Gly Ala Arg Phe Lys Leu Arg Ala Pro Ala Trp Gly Arg Cys Leu  
20 805 810 815  
Arg Thr His Ala Ser Gly Ala Gln Pro Asn Thr Asp Ile Ile Ala Ala  
820 825 830  
Leu Ser Glu Leu Pro Thr Glu Ala Trp Pro Met Met Gln Gly Ala Val  
835 840 845  
25 Asn Phe Ser Thr Leu  
850

## (2) INFORMATION FOR SEQ ID NO:271:

- 30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 857 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

35

- (ii) MOLECULE TYPE: peptide

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

40 Met Ala Glu Thr Met Asn Val Ala Thr Cys Thr His Gln Thr His His  
1 5 10 15  
Ala Ala Arg Ala Pro Gly Ala Thr Ser Ala Pro Gly Ala Ala Ser Gly  
20 25 30  
665



Asp Pro Leu Gly Ala Arg Arg Pro Ile Gly Asp Asp Glu Cys Glu Gln  
 35 40 45  
 Tyr Thr Ser Ser Val Ser Leu Ala Arg Met Leu Tyr Gly Gly Asp Leu  
 50 55 60  
 5 Ala Glu Trp Val Pro Arg Val His Pro Lys Thr Thr Ile Glu Arg Gln  
 65 70 75 80  
 Gln His Gly Pro Val Thr Phe Pro Asp Ala Ser Ala Pro Thr Ala Arg  
 85 90 95  
 Cys Val Thr Val Val Arg Ala Pro Met Gly Ser Gly Lys Thr Thr Ala  
 10 100 105 110  
 Leu Ile Arg Trp Leu Gly Glu Ala Ile His Ser Pro Asp Thr Ser Val  
 115 120 125  
 Leu Val Val Ser Cys Arg Arg Ser Phe Thr Gln Thr Leu Ala Thr Arg  
 130 135 140  
 15 Phe Ala Glu Ser Gly Leu Pro Asp Phe Val Thr Tyr Phe Ser Ser Thr  
 145 150 155 160  
 Asn Tyr Ile Met Asn Asp Arg Pro Phe His Arg Leu Ile Val Gln Val  
 165 170 175  
 Glu Ser Leu His Arg Val Gly Pro Asn Leu Leu Asn Asn Tyr Asp Val  
 180 185 190  
 20 Leu Val Leu Asp Glu Val Met Ser Thr Leu Gly Gln Lys Pro Thr Met  
 195 200 205  
 Gln Gln Leu Gly Arg Val Asp Ala Leu Met Leu Arg Leu Leu Arg Thr  
 210 215 220  
 25 Cys Pro Arg Ile Ile Ala Met Asp Ala Thr Ala Asn Ala Gln Leu Val  
 225 230 235 240  
 Asp Phe Leu Cys Ser Leu Arg Gly Glu Lys Asn Val His Val Val Ile  
 245 250 255  
 Gly Glu Tyr Ala Met Pro Gly Phe Ser Ala Arg Arg Cys Leu Phe Leu  
 260 265 270  
 30 Pro Arg Leu Gly Pro Glu Val Leu Gln Ala Ala Leu Arg Pro Pro Gly  
 275 280 285  
 Pro Ala Gly Gly Ala Pro Pro Pro Asp Ala Pro Pro Asp Ala Thr Phe  
 290 295 300  
 35 Phe Gly Glu Leu Glu Ala Arg Leu Ala Gly Gly Asp Asn Val Cys Ile  
 305 310 315 320  
 Phe Ser Ser Thr Val Ser Phe Ala Glu Val Val Ala Arg Phe Cys Arg  
 325 330 335  
 Gln Phe Thr Asp Arg Val Leu Leu Leu His Ser Leu Thr Pro Pro Gly  
 340 345 350  
 40 Asp Val Thr Thr Trp Gly Arg Tyr Arg Val Val Ile Tyr Thr Thr Val  
 355 360 365  
 Val Thr Val Gly Leu Ser Phe Asp Pro Pro His Phe Asp Ser Met Phe

	370		375		380	
	Ala Tyr Val Lys Pro Met Asn Tyr Gly Pro Asp Met Val Ser Val Tyr					
	385		390		395	400
	Gln Ser Leu Gly Arg Val Arg Thr Leu Arg Lys Gly Glu Leu Leu Ile					
5		405		410		415
	Tyr Met Asp Gly Ser Gly Ala Arg Ser Glu Pro Val Phe Thr Pro Met					
		420		425		430
	Leu Leu Asn His Val Val Ser Ala Ser Gly Gln Trp Pro Ala Gln Phe					
		435		440		445
10	Ser Gln Val Thr Asn Leu Leu Cys Arg Arg Phe Lys Gly Arg Cys Asp					
		450		455		460
	Ala Ser His Ala Asp Ala Ala Gln Arg Ser Arg Ile Tyr Ser Lys Phe					
		465		470		475
	Arg Tyr Lys His Tyr Phe Glu Arg Cys Thr Leu Ala Cys Leu Ala Asp					
15		485		490		495
	Ser Leu Asn Ile Leu His Met Leu Leu Thr Leu Asn Cys Met His Val					
		500		505		510
	Arg Phe Trp Gly His Asp Ala Ala Leu Thr Pro Arg Asn Phe Cys Leu					
		515		520		525
20	Phe Leu Arg Gly Ile His Phe Asp Ala Leu Arg Ala Gln Arg Asp Leu					
		530		535		540
	Arg Glu Leu Arg Cys Gln Asp Pro Asp Thr Ser Leu Ser Ala Gln Ala					
		545		550		555
	Ala Glu Thr Glu Glu Val Gly Leu Phe Val Glu Lys Tyr Leu Arg Pro					
25		565		570		575
	Asp Val Ala Pro Ala Glu Val Val Met Arg Gln Ser Leu Val Gly Arg					
		580		585		590
	Thr Arg Phe Ile Tyr Leu Val Leu Leu Glu Ala Cys Leu Arg Val Pro					
		595		600		605
30	Met Ala Ala His Ser Ser Ala Ile Phe Arg Arg Leu Tyr Asp His Tyr					
		610		615		620
	Ala Thr Gly Val Ile Pro Thr Ile Asn Ala Ala Gly Glu Leu Glu Leu					
		625		630		635
	Val His Pro Thr Leu Asn Val Ala Pro Val Trp Glu Leu Phe Arg Leu					
35		645		650		655
	Cys Ser Thr Met Ala Ala Cys Leu Gln Trp Asp Ser Met Ala Gly Gly					
		660		665		670
	Ser Gly Arg Thr Phe Ser Pro Glu Asp Val Leu Glu Leu Leu Asn Pro					
		675		680		685
40	His Tyr Asp Arg Tyr Met Gln Leu Val Phe Glu Leu Gly His Cys Asn					
		690		695		700
	Val Thr Asp Gly Pro Leu Leu Ser Glu Asp Ala Val Lys Arg Val Ala					
		705		710		715
						720

Asp Ala Leu Ser Gly Cys Pro Pro Arg Gly Ser Val Ser Glu Thr Glu  
                     725                    730                    735  
 His Ala Leu Ser Leu Phe Lys Ile Ile Trp Gly Glu Leu Phe Gly Val  
                     740                    745                    750  
 5 Gln Leu Ala Lys Ser Thr Gln Thr Phe Pro Gly Ala Gly Arg Val Lys  
                     755                    760                    765  
 Asn Leu Thr Lys Arg Ala Ile Val Glu Leu Leu Asp Ala His Arg Ile  
                     770                    775                    780  
 Asp His Ser Ala Cys Arg Thr Gln Leu Tyr Ala Leu Leu Met Ala His  
 10 785                    790                    795                    800  
 Lys Arg Glu Phe Ala Gly Ala Arg Phe Lys Leu Arg Ala Pro Ala Trp  
                     805                    810                    815  
 Gly Arg Cys Leu Arg Thr His Ala Ser Gly Ala Gln Pro Asn Thr Asp  
                     820                    825                    830  
 15 Ile Ile Ala Ala Leu Ser Glu Leu Pro Thr Glu Ala Trp Pro Met Met  
                     835                    840                    845  
 Gln Gly Ala Val Asn Phe Ser Thr Leu  
                     850                    855

20 (2) INFORMATION FOR SEQ ID NO:272:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1370 amino acids  
 (B) TYPE: amino acid  
 25 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

Met Glu Pro Ala Asn Pro Pro Arg Asn Pro Met Ala Ala Pro Ala Arg  
   1                    5                    10                    15  
 Asp Pro Pro Gly Tyr Arg Tyr Ala Ala Met Val Pro Thr Gly Ser  
 35                    20                    25                    30  
 Ile Leu Ser Thr Ile Glu Val Ala Ser His Arg Arg Leu Phe Asp Phe  
                     35                    40                    45  
 Phe Ala Arg Val Arg Ser Asp Glu Asn Ser Leu Tyr Asp Val Glu Phe  
                     50                    55                    60  
 40 Asp Ala Leu Leu Gly Ser Tyr Cys Asn Thr Leu Ser Leu Val Arg Phe  
                     65                    70                    75                    80  
 Leu Glu Leu Gly Leu Ser Val Ala Cys Val Cys Thr Lys Phe Pro Glu  
                     85                    90                    95

Leu Ala Tyr Met Asn Glu Gly Arg Val Gln Phe Glu Val His Gln Pro  
 100 105 110  
 Leu Ile Ala Arg Asp Gly Pro His Pro Val Glu Gln Pro Val His Asn  
 115 120 125  
 5 Tyr Met Thr Lys Val Ile Asp Arg Arg Ala Leu Asn Ala Ala Phe Ser  
 130 135 140  
 Leu Ala Thr Glu Ala Ile Ala Leu Leu Thr Gly Glu Ala Leu Asp Gly  
 145 150 155 160  
 Thr Gly Ile Ser Leu His Arg Gln Leu Arg Ala Ile Gln Gln Leu Ala  
 10 165 170 175  
 Arg Asn Val Gln Ala Val Leu Gly Ala Phe Glu Arg Gly Thr Ala Asp  
 180 185 190  
 Gln Met Leu His Val Leu Leu Glu Lys Ala Pro Pro Leu Ala Leu Leu  
 195 200 205  
 15 Leu Pro Met Gln Arg Tyr Leu Asp Asn Gly Arg Leu Ala Thr Arg Val  
 210 215 220  
 Ala Arg Ala Thr Leu Val Ala Glu Leu Lys Arg Ser Phe Cys Asp Thr  
 225 230 235 240  
 Ser Phe Phe Leu Gly Lys Ala Gly His Arg Arg Glu Ala Ile Glu Ala  
 20 245 250 255  
 Trp Leu Val Asp Leu Thr Thr Ala Thr Gln Pro Ser Val Ala Val Pro  
 260 265 270  
 Arg Leu Thr His Ala Asp Thr Arg Gly Arg Pro Val Asp Gly Val Leu  
 275 280 285  
 25 Val Thr Thr Ala Ala Ile Lys Gln Arg Leu Leu Gln Ser Phe Leu Lys  
 290 295 300  
 Val Glu Asp Thr Glu Ala Asp Val Pro Val Thr Tyr Gly Glu Met Val  
 305 310 315 320  
 Leu Asn Gly Ala Asn Leu Val Thr Ala Leu Val Met Gly Lys Ala Val  
 30 325 330 335  
 Arg Ser Leu Asp Asp Val Gly Arg His Leu Leu Glu Met Gln Glu Glu  
 340 345 350  
 Gln Leu Glu Ala Asn Arg Glu Thr Leu Asp Glu Leu Glu Ser Ala Pro  
 355 360 365  
 35 Gln Thr Thr Arg Val Arg Ala Asp Leu Val Ala Ile Gly Asp Arg Leu  
 370 375 380  
 Val Phe Leu Glu Ala Leu Glu Lys Arg Ile Tyr Ala Ala Thr Asn Val  
 385 390 395 400  
 Pro Tyr Pro Leu Val Gly Ala Met Asp Leu Thr Phe Val Leu Pro Leu  
 40 405 410 415  
 Gly Leu Phe Asn Pro Ala Met Glu Arg Phe Ala Ala His Ala Gly Asp  
 420 425 430  
 Leu Val Pro Ala Pro Gly His Pro Glu Pro Arg Ala Phe Pro Pro Arg

	435		440		445	
	Gln Leu Phe Phe Trp Gly Lys Asp His Gln Val Leu Arg Leu Ser Met					
	450		455		460	
	Glu Asn Ala Val Gly Thr Val Cys His Pro Ser Leu Met Asn Ile Asp					
5	465		470		475	480
	Ala Ala Val Gly Gly Val Asn His Asp Pro Val Glu Ala Ala Asn Pro					
		485		490		495
	Tyr Gly Ala Tyr Val Ala Ala Pro Ala Gly Pro Gly Ala Asp Met Gln					
		500		505		510
10	Gln Arg Phe Leu Asn Ala Trp Arg Gln Arg Leu Ala His Gly Arg Val					
		515		520		525
	Arg Trp Val Ala Glu Cys Gln Met Thr Ala Glu Gln Phe Met Gln Pro					
		530		535		540
	Asp Asn Ala Asn Leu Ala Leu Glu Leu His Pro Ala Phe Asp Phe Phe					
15	545		550		555	560
	Ala Gly Val Ala Asp Val Glu Leu Pro Gly Gly Glu Val Pro Pro Ala					
		565		570		575
	Gly Pro Gly Ala Ile Gln Ala Thr Trp Arg Val Val Asn Gly Asn Leu					
		580		585		590
20	Pro Leu Ala Leu Cys Pro Val Ala Phe Arg Asp Arg Leu Glu Leu Gly					
		595		600		605
	Val Gly Arg His Ala Met Ala Pro Ala Thr Ile Ala Ala Val Arg Gly					
		610		615		620
	Ala Phe Glu Asp Arg Ser Tyr Pro Ala Val Phe Tyr Leu Leu Gln Ala					
25	625		630		635	640
	Ala Ile His Gly Ser Glu His Val Phe Cys Ala Arg Leu Val Thr Gln					
		645		650		655
	Cys Ile Thr Ser Tyr Trp Asn Asn Thr Arg Cys Ala Ala Phe Val Asn					
		660		665		670
30	Asp Tyr Ser Leu Val Ser Tyr Ile Val Thr Tyr Leu Gly Gly Asp Leu					
		675		680		685
	Pro Glu Glu Cys Met Ala Val Tyr Arg Asp Leu Val Ala His Val Glu					
		690		695		700
	Ala Gln Leu Val Asp Asp Phe Thr Leu Pro Gly Pro Glu Leu Gly Gly					
35	705		710		715	720
	Gln Ala Gln Ala Glu Leu Asn His Leu Met Arg Asp Pro Ala Leu Leu					
		725		730		735
	Pro Pro Leu Val Trp Asp Cys Asp Gly Leu Met Arg His Ala Ala Leu					
		740		745		750
40	Asp Arg His Arg Asp Cys Arg Ile Asp Ala Gly Gly His Glu Pro Val					
		755		760		765
	Tyr Ala Ala Ala Cys Asn Val Ala Thr Ala Asp Phe Asn Arg Asn Asp					
		770		775		780

Gly Arg Leu Leu His Asn Thr Gln Ala Arg Ala Ala Asp Ala Ala Asp  
 785 790 795 800  
 Asp Arg Pro His Arg Pro Ala Asp Trp Thr Val His His Lys Ile Tyr  
 805 810 815  
 5 Tyr Tyr Val Leu Val Pro Ala Phe Ser Arg Gly Arg Cys Cys Thr Ala  
 820 825 830  
 Gly Val Arg Phe Asp Arg Val Tyr Ala Thr Leu Gln Asn Met Val Val  
 835 840 845  
 Pro Glu Ile Ala Pro Gly Glu Glu Cys Pro Ser Asp Pro Val Thr Asp  
 10 850 855 860  
 Pro Ala His Pro Leu His Pro Ala Asn Leu Val Ala Asn Thr Val Asn  
 865 870 875 880  
 Ala Met Phe His Asn Gly Arg Val Val Val Asp Gly Pro Ala Met Leu  
 885 890 895  
 15 Thr Leu Gln Val Leu Ala His Asn Met Ala Glu Arg Thr Thr Ala Leu  
 900 905 910  
 Leu Cys Ser Ala Ala Pro Asp Ala Gly Ala Asn Thr Ala Ser Thr Ala  
 915 920 925  
 Asn Met Arg Ile Phe Asp Gly Ala Leu His Ala Gly Val Leu Leu Met  
 20 930 935 940  
 Ala Pro Gln His Leu Asp His Thr Ile Gln Asn Gly Glu Tyr Phe Tyr  
 945 950 955 960  
 Val Leu Pro Val His Ala Leu Phe Ala Gly Ala Asp His Val Ala Asn  
 965 970 975  
 25 Ala Pro Asn Phe Pro Pro Ala Leu Arg Asp Leu Ala Arg His Val Pro  
 980 985 990  
 Leu Val Pro Pro Ala Leu Gly Ala Asn Tyr Phe Ser Ser Ile Arg Gln  
 995 1000 1005  
 Pro Val Val Gln His Ala Arg Glu Ser Ala Ala Gly Glu Asn Ala Leu  
 30 1010 1015 1020  
 Thr Tyr Ala Leu Met Ala Gly Tyr Phe Lys Met Ser Pro Val Tyr His  
 1025 1030 1035 104  
 Gln Leu Lys Thr Gly Leu His Pro Gly Phe Gly Phe Thr Val Val Arg  
 1045 1050 1055  
 35 Gln Asp Arg Phe Val Thr Glu Asn Val Leu Phe Ser Ala Ser Glu Ala  
 1060 1065 1070  
 Tyr Phe Leu Gly Gln Leu Gln Val Ala Arg His Glu Thr Gly Gly Gly  
 1075 1080 1085  
 Val Ser Phe Thr Leu Thr Gln Pro Arg Gly Asn Val Asp Leu Gly Val  
 40 1090 1095 1100  
 Gly Tyr Thr Ala Val Ala Ala Thr Ala Thr Val Arg Asn Pro Val Thr  
 1105 1110 1115 112  
 Asp Met Gly Asn Leu Pro Gln Asn Phe Tyr Leu Gly Arg Gly Ala Pro

	1125	1130	1135
	Pro Leu Leu Asp Asn Ala Ala Ala Val Tyr Leu Arg Asn Ala Val Val		
	1140	1145	1150
	Ala Gly Asn Arg Leu Gly Pro Ala Gln Pro Leu Pro Val Phe Gly Cys		
5	1155	1160	1165
	Ala Gln Val Pro Arg Arg Ala Gly Met Asp His Gly Gln Asp Ala Val		
	1170	1175	1180
	Cys Glu Phe Ile Ala Thr Pro Val Ala Thr Asp Ile Asn Tyr Phe Arg		
	1185	1190	1195
10	Arg Pro Cys Asn Pro Arg Gly Arg Ala Ala Gly Gly Val Tyr Ala Gly		120
	1205	1210	1215
	Asp Lys Glu Gly Asp Val Ile Ala Leu Met Tyr Asp His Gly Gln Ser		
	1220	1225	1230
	Asp Pro Ala Arg Pro Phe Ala Ala Thr Ala Asn Pro Trp Ala Ser Gln		
15	1235	1240	1245
	Arg Phe Ser Tyr Gly Asp Leu Leu Tyr Asn Gly Ala Tyr His Leu Asn		
	1250	1255	1260
	Gly Asp Val Leu Ser Pro Cys Phe Lys Phe Phe Thr Ala Ala Asp Ile		
	1265	1270	1275
20	Thr Ala Lys His Arg Cys Leu Glu Arg Leu Ile Val Glu Thr Gly Ser		128
	1285	1290	1295
	Ala Val Ser Thr Ala Thr Ala Ala Ser Asp Val Gln Phe Lys Arg Pro		
	1300	1305	1310
	Pro Gly Cys Arg Glu Leu Val Glu Asp Pro Cys Gly Leu Phe Gln Glu		
25	1315	1320	1325
	Ala Tyr Pro Ile Thr Cys Ala Ser Asp Pro Ala Leu Leu Arg Ser Ala		
	1330	1335	1340
	Arg Asp Gly Glu Ala His Ala Arg Glu Thr His Phe Thr Gln Tyr Leu		
	1345	1350	1355
30	Ile Tyr Asp Asp Leu Lys Gly Leu Ser Leu		136
	1365	1370	

## (2) INFORMATION FOR SEQ ID NO:273:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1360 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

40

- (ii) MOLECULE TYPE: peptide

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:273:

Met Ala Ala Pro Ala Arg Asp Pro Pro Gly Tyr Arg Tyr Ala Ala Ala  
 1 5 10 15  
 Met Val Pro Thr Gly Ser Ile Leu Ser Thr Ile Glu Val Ala Ser His  
 5 20 25 30  
 Arg Arg Leu Phe Asp Phe Phe Ala Arg Val Arg Ser Asp Glu Asn Ser  
 35 40 45  
 Leu Tyr Asp Val Glu Phe Asp Ala Leu Leu Gly Ser Tyr Cys Asn Thr  
 50 55 60  
 10 Leu Ser Leu Val Arg Phe Leu Glu Leu Gly Leu Ser Val Ala Cys Val  
 65 70 75 80  
 Cys Thr Lys Phe Pro Glu Leu Ala Tyr Met Asn Glu Gly Arg Val Gln  
 85 90 95  
 Phe Glu Val His Gln Pro Leu Ile Ala Arg Asp Gly Pro His Pro Val  
 15 100 105 110  
 Glu Gln Pro Val His Asn Tyr Met Thr Lys Val Ile Asp Arg Arg Ala  
 115 120 125  
 Leu Asn Ala Ala Phe Ser Leu Ala Thr Glu Ala Ile Ala Leu Leu Thr  
 130 135 140  
 20 Gly Glu Ala Leu Asp Gly Thr Gly Ile Ser Leu His Arg Gln Leu Arg  
 145 150 155 160  
 Ala Ile Gln Gln Leu Ala Arg Asn Val Gln Ala Val Leu Gly Ala Phe  
 165 170 175  
 Glu Arg Gly Thr Ala Asp Gln Met Leu His Val Leu Leu Glu Lys Ala  
 25 180 185 190  
 Pro Pro Leu Ala Leu Leu Leu Pro Met Gln Arg Tyr Leu Asp Asn Gly  
 195 200 205  
 Arg Leu Ala Thr Arg Val Ala Arg Ala Thr Leu Val Ala Glu Leu Lys  
 210 215 220  
 30 Arg Ser Phe Cys Asp Thr Ser Phe Phe Leu Gly Lys Ala Gly His Arg  
 225 230 235 240  
 Arg Glu Ala Ile Glu Ala Trp Leu Val Asp Leu Thr Thr Ala Thr Gln  
 245 250 255  
 Pro Ser Val Ala Val Pro Arg Leu Thr His Ala Asp Thr Arg Gly Arg  
 35 260 265 270  
 Pro Val Asp Gly Val Leu Val Thr Thr Ala Ala Ile Lys Gln Arg Leu  
 275 280 285  
 Leu Gln Ser Phe Leu Lys Val Glu Asp Thr Glu Ala Asp Val Pro Val  
 290 295 300  
 40 Thr Tyr Gly Glu Met Val Leu Asn Gly Ala Asn Leu Val Thr Ala Leu  
 305 310 315 320  
 Val Met Gly Lys Ala Val Arg Ser Leu Asp Asp Val Gly Arg His Leu  
 325 330 335



Leu Glu Met Gln Glu Glu Gln Leu Glu Ala Asn Arg Glu Thr Leu Asp  
 340 345 350  
 Glu Leu Glu Ser Ala Pro Gln Thr Thr Arg Val Arg Ala Asp Leu Val  
 355 360 365  
 5 Ala Ile Gly Asp Arg Leu Val Phe Leu Glu Ala Leu Glu Lys Arg Ile  
 370 375 380  
 Tyr Ala Ala Thr Asn Val Pro Tyr Pro Leu Val Gly Ala Met Asp Leu  
 385 390 395 400  
 Thr Phe Val Leu Pro Leu Gly Leu Phe Asn Pro Ala Met Glu Arg Phe  
 10 405 410 415  
 Ala Ala His Ala Gly Asp Leu Val Pro Ala Pro Gly His Pro Glu Pro  
 420 425 430  
 Arg Ala Phe Pro Pro Arg Gln Leu Phe Phe Trp Gly Lys Asp His Gln  
 435 440 445  
 15 Val Leu Arg Leu Ser Met Glu Asn Ala Val Gly Thr Val Cys His Pro  
 450 455 460  
 Ser Leu Met Asn Ile Asp Ala Ala Val Gly Gly Val Asn His Asp Pro  
 465 470 475 480  
 Val Glu Ala Ala Asn Pro Tyr Gly Ala Tyr Val Ala Ala Pro Ala Gly  
 20 485 490 495  
 Pro Gly Ala Asp Met Gln Gln Arg Phe Leu Asn Ala Trp Arg Gln Arg  
 500 505 510  
 Leu Ala His Gly Arg Val Arg Trp Val Ala Glu Cys Gln Met Thr Ala  
 515 520 525  
 25 Glu Gln Phe Met Gln Pro Asp Asn Ala Asn Leu Ala Leu Glu Leu His  
 530 535 540  
 Pro Ala Phe Asp Phe Phe Ala Gly Val Ala Asp Val Glu Leu Pro Gly  
 545 550 555 560  
 Gly Glu Val Pro Pro Ala Gly Pro Gly Ala Ile Gln Ala Thr Trp Arg  
 30 565 570 575  
 Val Val Asn Gly Asn Leu Pro Leu Ala Leu Cys Pro Val Ala Phe Arg  
 580 585 590  
 Asp Arg Leu Glu Leu Gly Val Gly Arg His Ala Met Ala Pro Ala Thr  
 595 600 605  
 35 Ile Ala Ala Val Arg Gly Ala Phe Glu Asp Arg Ser Tyr Pro Ala Val  
 610 615 620  
 Phe Tyr Leu Leu Gln Ala Ala Ile His Gly Ser Glu His Val Phe Cys  
 625 630 635 640  
 Ala Arg Leu Val Thr Gln Cys Ile Thr Ser Tyr Trp Asn Asn Thr Arg  
 40 645 650 655  
 Cys Ala Ala Phe Val Asn Asp Tyr Ser Leu Val Ser Tyr Ile Val Thr  
 660 665 670  
 Tyr Leu Gly Gly Asp Leu Pro Glu Glu Cys Met Ala Val Tyr Arg Asp  
 674

	675	680	685
	Leu Val Ala His Val Glu Ala Gln Leu Val Asp Asp Phe Thr Leu Pro		
	690	695	700
	Gly Pro Glu Leu Gly Gly Gln Ala Gln Ala Glu Leu Asn His Leu Met		
5	705	710	715
	Arg Asp Pro Ala Leu Leu Pro Pro Leu Val Trp Asp Cys Asp Gly Leu		
	725	730	735
	Met Arg His Ala Ala Leu Asp Arg His Arg Asp Cys Arg Ile Asp Ala		
	740	745	750
10	Gly Gly His Glu Pro Val Tyr Ala Ala Ala Cys Asn Val Ala Thr Ala		
	755	760	765
	Asp Phe Asn Arg Asn Asp Gly Arg Leu Leu His Asn Thr Gln Ala Arg		
	770	775	780
	Ala Ala Asp Ala Ala Asp Asp Arg Pro His Arg Pro Ala Asp Trp Thr		
15	785	790	795
	Val His His Lys Ile Tyr Tyr Tyr Val Leu Val Pro Ala Phe Ser Arg		
	805	810	815
	Gly Arg Cys Cys Thr Ala Gly Val Arg Phe Asp Arg Val Tyr Ala Thr		
	820	825	830
20	Leu Gln Asn Met Val Val Pro Glu Ile Ala Pro Gly Glu Glu Cys Pro		
	835	840	845
	Ser Asp Pro Val Thr Asp Pro Ala His Pro Leu His Pro Ala Asn Leu		
	850	855	860
	Val Ala Asn Thr Val Asn Ala Met Phe His Asn Gly Arg Val Val Val		
25	865	870	875
	Asp Gly Pro Ala Met Leu Thr Leu Gln Val Leu Ala His Asn Met Ala		
	885	890	895
	Glu Arg Thr Thr Ala Leu Leu Cys Ser Ala Ala Pro Asp Ala Gly Ala		
	900	905	910
30	Asn Thr Ala Ser Thr Ala Asn Met Arg Ile Phe Asp Gly Ala Leu His		
	915	920	925
	Ala Gly Val Leu Leu Met Ala Pro Gln His Leu Asp His Thr Ile Gln		
	930	935	940
	Asn Gly Glu Tyr Phe Tyr Val Leu Pro Val His Ala Leu Phe Ala Gly		
35	945	950	955
	Ala Asp His Val Ala Asn Ala Pro Asn Phe Pro Pro Ala Leu Arg Asp		
	965	970	975
	Leu Ala Arg His Val Pro Leu Val Pro Pro Ala Leu Gly Ala Asn Tyr		
	980	985	990
40	Phe Ser Ser Ile Arg Gln Pro Val Val Gln His Ala Arg Glu Ser Ala		
	995	1000	1005
	Ala Gly Glu Asn Ala Leu Thr Tyr Ala Leu Met Ala Gly Tyr Phe Lys		
	1010	1015	1020

Met Ser Pro Val Tyr His Gln Leu Lys Thr Gly Leu His Pro Gly Phe  
 1025 1030 1035 104  
 Gly Phe Thr Val Val Arg Gln Asp Arg Phe Val Thr Glu Asn Val Leu  
 1045 1050 1055  
 5 Phe Ser Ala Ser Glu Ala Tyr Phe Leu Gly Gln Leu Gln Val Ala Arg  
 1060 1065 1070  
 His Glu Thr Gly Gly Gly Val Ser Phe Thr Leu Thr Gln Pro Arg Gly  
 1075 1080 1085  
 Asn Val Asp Leu Gly Val Gly Tyr Thr Ala Val Ala Ala Thr Ala Thr  
 10 1090 1095 1100  
 Val Arg Asn Pro Val Thr Asp Met Gly Asn Leu Pro Gln Asn Phe Tyr  
 1105 1110 1115 112  
 Leu Gly Arg Gly Ala Pro Pro Leu Leu Asp Asn Ala Ala Ala Val Tyr  
 1125 1130 1135  
 15 Leu Arg Asn Ala Val Val Ala Gly Asn Arg Leu Gly Pro Ala Gln Pro  
 1140 1145 1150  
 Leu Pro Val Phe Gly Cys Ala Gln Val Pro Arg Arg Ala Gly Met Asp  
 1155 1160 1165  
 His Gly Gln Asp Ala Val Cys Glu Phe Ile Ala Thr Pro Val Ala Thr  
 20 1170 1175 1180  
 Asp Ile Asn Tyr Phe Arg Arg Pro Cys Asn Pro Arg Gly Arg Ala Ala  
 1185 1190 1195 120  
 Gly Gly Val Tyr Ala Gly Asp Lys Glu Gly Asp Val Ile Ala Leu Met  
 1205 1210 1215  
 25 Tyr Asp His Gly Gln Ser Asp Pro Ala Arg Pro Phe Ala Ala Thr Ala  
 1220 1225 1230  
 Asn Pro Trp Ala Ser Gln Arg Phe Ser Tyr Gly Asp Leu Leu Tyr Asn  
 1235 1240 1245  
 Gly Ala Tyr His Leu Asn Gly Asp Val Leu Ser Pro Cys Phe Lys Phe  
 30 1250 1255 1260  
 Phe Thr Ala Ala Asp Ile Thr Ala Lys His Arg Cys Leu Glu Arg Leu  
 1265 1270 1275 128  
 Ile Val Glu Thr Gly Ser Ala Val Ser Thr Ala Thr Ala Ala Ser Asp  
 1285 1290 1295  
 35 Val Gln Phe Lys Arg Pro Pro Gly Cys Arg Glu Leu Val Glu Asp Pro  
 1300 1305 1310  
 Cys Gly Leu Phe Gln Glu Ala Tyr Pro Ile Thr Cys Ala Ser Asp Pro  
 1315 1320 1325  
 Ala Leu Leu Arg Ser Ala Arg Asp Gly Glu Ala His Ala Arg Glu Thr  
 40 1330 1335 1340  
 His Phe Thr Gln Tyr Leu Ile Tyr Asp Asp Leu Lys Gly Leu Ser Leu  
 1345 1350 1355 136

## (2) INFORMATION FOR SEQ ID NO:274:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 604 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

Met Arg Pro Glu Leu Ser Leu Lys Gly Arg Pro Cys Val Thr Glu Ala  
 1 5 10 15  
 15 Val Val Cys Pro Ser Thr Asp Ala Ala Ile His Ser Gly Gly Ser Ser  
 20 25 30  
 Ser Val Arg Pro Gln Pro Tyr Ala Arg Ala Ala Arg Ala Arg Ala Thr  
 35 40 45  
 His Gly Ser Arg Ser Arg His Arg Gln Pro Leu Leu Pro Pro Pro Ser  
 20 50 55 60  
 Ser His His Pro Thr Ile Pro Pro Pro Pro Ser Pro Pro Arg Gly Ser  
 65 70 75 80  
 Pro Ala Met Glu Leu Ser Tyr Ala Thr Thr Leu His His Arg Asp Val  
 85 90 95  
 25 Val Phe Tyr Val Thr Ala Asp Arg Asn Arg Ala Tyr Phe Val Cys Gly  
 100 105 110  
 Gly Ser Val Tyr Ser Val Gly Arg Pro Arg Asp Ser Gln Pro Gly Glu  
 115 120 125  
 Ile Ala Lys Phe Gly Leu Val Val Arg Gly Thr Gly Pro Lys Asp Arg  
 30 130 135 140  
 Met Val Ala Asn Tyr Val Arg Ser Glu Leu Arg Gln Arg Gly Leu Arg  
 145 150 155 160  
 Asp Val Arg Pro Val Gly Glu Asp Glu Val Phe Leu Asp Ser Val Cys  
 165 170 175  
 35 Leu Leu Asn Pro Asn Val Ser Ser Asp Val Ile Asn Thr Asn Asp Val  
 180 185 190  
 Glu Val Leu Asp Glu Cys Leu Ala Glu Tyr Cys Thr Ser Leu Arg Thr  
 195 200 205  
 Ser Pro Gly Val Leu Val Thr Gly Val Arg Val Arg Ala Arg Asp Arg  
 40 210 215 220  
 Val Ile Glu Leu Phe Glu His Pro Ala Ile Val Asn Ile Ser Ser Arg  
 225 230 235 240  
 Phe Ala Tyr Thr Pro Ser Pro Tyr Val Phe Ala Gln Ala His Leu Pro

						245					250					255
	Arg	Leu	Pro	Ser	Ser	Leu	Glu	Pro	Leu	Val	Ser	Gly	Leu	Phe	Asp	Gly
				260					265					270		
5	Ile	Pro	Ala	Pro	Arg	Gln	Pro	Leu	Asp	Ala	Arg	Asp	Arg	Arg	Thr	Asp
			275					280					285			
	Val	Val	Ile	Thr	Gly	Thr	Arg	Ala	Pro	Arg	Pro	Met	Ala	Gly	Thr	Gly
		290					295					300				
	Ala	Gly	Gly	Ala	Gly	Ala	Lys	Arg	Ala	Thr	Val	Ser	Glu	Phe	Val	Gln
	305					310					315					320
10	Val	Lys	His	Ile	Asp	Arg	Val	Val	Ser	Pro	Ser	Val	Ser	Ser	Ala	Pro
				325						330					335	
	Pro	Pro	Ser	Ala	Pro	Asp	Ala	Ser	Leu	Pro	Pro	Pro	Gly	Leu	Gln	Glu
				340					345					350		
	Ala	Ala	Pro	Pro	Gly	Pro	Pro	Leu	Arg	Glu	Leu	Trp	Trp	Val	Phe	Tyr
15			355					360					365			
	Ala	Gly	Asp	Arg	Ala	Leu	Glu	Glu	Pro	His	Ala	Glu	Ser	Gly	Leu	Thr
	370						375					380				
	Arg	Glu	Glu	Val	Arg	Ala	Val	His	Gly	Phe	Arg	Glu	Gln	Ala	Trp	Lys
	385				390					395						400
20	Leu	Phe	Gly	Ser	Val	Gly	Ala	Pro	Arg	Ala	Phe	Leu	Gly	Ala	Ala	Leu
				405						410					415	
	Ser	Pro	Thr	Gln	Lys	Leu	Ala	Val	Tyr	Tyr	Tyr	Leu	Ile	His	Arg	Glu
				420					425					430		
	Arg	Arg	Met	Ser	Pro	Phe	Pro	Ala	Leu	Val	Arg	Leu	Val	Gly	Arg	Tyr
25			435					440					445			
	Ile	Gln	Arg	His	Gly	Val	Pro	Ala	Pro	Asp	Glu	Pro	Thr	Leu	Ala	Asp
	450						455					460				
	Ala	Met	Asn	Gly	Leu	Phe	Arg	Asp	Ala	Ala	Gly	Thr	Val	Ala	Glu	Gln
	465					470					475					480
30	Leu	Leu	Met	Phe	Asp	Leu	Leu	Pro	Pro	Lys	Asp	Val	Pro	Val	Gly	Ser
				485						490					495	
	Asp	Ala	Arg	Ala	Asp	Ser	Ala	Ala	Leu	Leu	Arg	Phe	Val	Asp	Ser	Gln
			500						505					510		
	Arg	Leu	Thr	Pro	Gly	Gly	Ser	Val	Ser	Pro	Glu	His	Val	Met	Tyr	Leu
35			515					520					525			

Val Cys Leu Ala Arg Ala Gln His Gly Gln Ser Val  
 595 600

(2) INFORMATION FOR SEQ ID NO:275:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 522 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

10

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

15

	Met	Glu	Leu	Ser	Tyr	Ala	Thr	Thr	Leu	His	His	Arg	Asp	Val	Val	Phe
	1				5				10					15		
	Tyr	Val	Thr	Ala	Asp	Arg	Asn	Arg	Ala	Tyr	Phe	Val	Cys	Gly	Gly	Ser
				20					25					30		
20	Val	Tyr	Ser	Val	Gly	Arg	Pro	Arg	Asp	Ser	Gln	Pro	Gly	Glu	Ile	Ala
			35					40					45			
	Lys	Phe	Gly	Leu	Val	Val	Arg	Gly	Thr	Gly	Pro	Lys	Asp	Arg	Met	Val
	50						55					60				
	Ala	Asn	Tyr	Val	Arg	Ser	Glu	Leu	Arg	Gln	Arg	Gly	Leu	Arg	Asp	Val
25	65				70					75				80		
	Arg	Pro	Val	Gly	Glu	Asp	Glu	Val	Phe	Leu	Asp	Ser	Val	Cys	Leu	Leu
				85					90					95		
	Asn	Pro	Asn	Val	Ser	Ser	Asp	Val	Ile	Asn	Thr	Asn	Asp	Val	Glu	Val
				100					105					110		
30	Leu	Asp	Glu	Cys	Leu	Ala	Glu	Tyr	Cys	Thr	Ser	Leu	Arg	Thr	Ser	Pro
			115					120					125			
	Gly	Val	Leu	Val	Thr	Gly	Val	Arg	Val	Arg	Ala	Arg	Asp	Arg	Val	Ile
			130				135					140				
	Glu	Leu	Phe	Glu	His	Pro	Ala	Ile	Val	Asn	Ile	Ser	Ser	Arg	Phe	Ala
35	145				150					155				160		
	Tyr	Thr	Pro	Ser	Pro	Tyr	Val	Phe	Ala	Gln	Ala	His	Leu	Pro	Arg	Leu
				165					170					175		
	Pro	Ser	Ser	Leu	Glu	Pro	Leu	Val	Ser	Gly	Leu	Phe	Asp	Gly	Ile	Pro
				180					185					190		
40	Ala	Pro	Arg	Gln	Pro	Leu	Asp	Ala	Arg	Asp	Arg	Arg	Thr	Asp	Val	Val
			195				200					205				
	Ile	Thr	Gly	Thr	Arg	Ala	Pro	Arg	Pro	Met	Ala	Gly	Thr	Gly	Ala	Gly
	210						215					220				

Gly Ala Gly Ala Lys Arg Ala Thr Val Ser Glu Phe Val Gln Val Lys  
 225 230 235 240  
 His Ile Asp Arg Val Val Ser Pro Ser Val Ser Ser Ala Pro Pro Pro  
 245 250 255  
 5 Ser Ala Pro Asp Ala Ser Leu Pro Pro Pro Gly Leu Gln Glu Ala Ala  
 260 265 270  
 Pro Pro Gly Pro Pro Leu Arg Glu Leu Trp Trp Val Phe Tyr Ala Gly  
 275 280 285  
 Asp Arg Ala Leu Glu Glu Pro His Ala Glu Ser Gly Leu Thr Arg Glu  
 10 290 295 300  
 Glu Val Arg Ala Val His Gly Phe Arg Glu Gln Ala Trp Lys Leu Phe  
 305 310 315 320  
 Gly Ser Val Gly Ala Pro Arg Ala Phe Leu Gly Ala Ala Leu Ser Pro  
 325 330 335  
 15 Thr Gln Lys Leu Ala Val Tyr Tyr Tyr Leu Ile His Arg Glu Arg Arg  
 340 345 350  
 Met Ser Pro Phe Pro Ala Leu Val Arg Leu Val Gly Arg Tyr Ile Gln  
 355 360 365  
 Arg His Gly Val Pro Ala Pro Asp Glu Pro Thr Leu Ala Asp Ala Met  
 20 370 375 380  
 Asn Gly Leu Phe Arg Asp Ala Ala Gly Thr Val Ala Glu Gln Leu Leu  
 385 390 395 400  
 Met Phe Asp Leu Leu Pro Pro Lys Asp Val Pro Val Gly Ser Asp Ala  
 405 410 415  
 25 Arg Ala Asp Ser Ala Ala Leu Leu Arg Phe Val Asp Ser Gln Arg Leu  
 420 425 430  
 Thr Pro Gly Gly Ser Val Ser Pro Glu His Val Met Tyr Leu Gly Ala  
 435 440 445  
 Phe Leu Gly Val Leu Tyr Ala Gly His Gly Arg Leu Ala Ala Ala Thr  
 30 450 455 460  
 His Thr Ala Arg Leu Thr Gly Val Thr Ser Leu Val Leu Thr Val Gly  
 465 470 475 480  
 Asp Val Asp Arg Met Ser Ala Phe Asp Arg Gly Pro Ala Gly Ala Ala  
 485 490 495  
 35 Gly Arg Thr Arg Thr Ala Gly Tyr Leu Asp Ala Leu Leu Thr Val Cys  
 500 505 510  
 Leu Ala Arg Ala Gln His Gly Gln Ser Val  
 515 520

40 (2) INFORMATION FOR SEQ ID NO:276:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 602 amino acids

680

(B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

	Met	Thr	Ala	Ala	Ala	Leu	Tyr	Gly	Gly	Ala	Lys	Tyr	Arg	Pro	Gly	Thr
10	1				5					10					15	
	Leu	Arg	Asn	Pro	Gly	Arg	Val	Ala	Ser	Thr	Pro	Arg	Arg	Arg	Gly	Val
				20				25						30		
	Leu	Tyr	Gly	Ala	Leu	Cys	Pro	Gly	Ile	Pro	Phe	Val	Gly	Ser	Gly	Pro
		35					40					45				
15	Gly	Ala	Val	Gly	Trp	Glu	Cys	Val	Cys	Val	Gly	Gly	Gly	Arg	Arg	Asp
	50					55				60						
	Gly	Gly	Pro	Asp	Gln	Val	Tyr	Arg	Gly	Arg	Ser	Val	Gly	Arg	Pro	Asn
	65				70					75					80	
	Arg	Pro	Phe	Lys	His	Leu	Arg	Met	His	Arg	Pro	Ser	Gln	Ser	Asp	Thr
20				85					90					95		
	Gly	Thr	His	Gln	Arg	Arg	Lys	Pro	Pro	Ser	Pro	Val	Arg	Val	Arg	Val
				100				105					110			
	Phe	Ser	Gly	Gly	Val	Phe	Phe	Leu	Ser	Ala	Leu	Leu	Pro	Pro	His	Leu
		115				120						125				
25	His	His	Pro	Pro	Pro	Thr	Trp	Leu	Ala	Ile	Gly	Gly	Lys	Thr	Met	Lys
	130					135					140					
	Thr	Lys	Pro	Leu	Pro	Thr	Ala	Pro	Met	Ala	Trp	Ala	Glu	Ser	Ala	Val
	145				150					155					160	
	Glu	Thr	Thr	Thr	Ser	Pro	Arg	Glu	Leu	Ala	Gly	His	Ala	Pro	Leu	Arg
30				165				170					175			
	Arg	Val	Leu	Arg	Pro	Pro	Ile	Ala	Arg	Arg	Asp	Gly	Pro	Val	Leu	Leu
				180				185					190			
	Gly	Asp	Arg	Ala	Pro	Arg	Arg	Thr	Ala	Ser	Thr	Met	Trp	Leu	Leu	Gly
		195				200						205				
35	Ile	Asp	Pro	Ala	Glu	Ser	Ser	Pro	Gly	Thr	Arg	Ala	Thr	Arg	Asp	Asp
	210					215						220				
	Thr	Glu	Gln	Ala	Val	Asp	Lys	Ile	Leu	Arg	Gly	Ala	Arg	Arg	Ala	Gly
	225				230					235					240	
	Gly	Leu	Thr	Val	Pro	Gly	Ala	Pro	Arg	Tyr	His	Leu	Thr	Arg	Gln	Val
40				245				250					255			
	Thr	Leu	Thr	Asp	Leu	Cys	Gln	Pro	Asn	Ala	Glu	Arg	Ala	Gly	Ala	Leu
				260				265					270			
	Leu	Leu	Ala	Leu	Arg	His	Pro	Thr	Asp	Leu	Pro	His	Leu	Ala	Arg	His



	275		280		285
	Arg Ala Pro Pro Gly Arg Gln Thr Glu Arg Leu Ala Glu Ala Trp Gly				
	290		295		300
	Gln Leu Leu Glu Ala Ser Ala Leu Gly Ser Gly Arg Ala Glu Ser Gly				
5	305		310		315
	Cys Ala Arg Ala Gly Leu Val Ser Phe Asn Phe Leu Val Ala Ala Cys				
		325		330	335
	Ala Ala Ala Tyr Asp Ala Arg Asp Ala Ala Glu Ala Val Arg Ala His				
		340		345	350
10	Ile Thr Thr Asn Tyr Gly Gly Thr Arg Ala Gly Ala Arg Leu Asp Arg				
		355		360	365
	Phe Ser Glu Cys Leu Arg Ala Met Val His Thr His Val Phe Phe Val				
		370		375	380
	Met Arg Phe Phe Gly Gly Leu Val Ser Trp Val Thr Gln Asp Glu Leu				
15	385		390		395
	Ala Ser Val Thr Ala Val Cys Ser Gly Pro Gln Glu Ala Thr His Thr				
		405		410	415
	Gly His Pro Gly Arg Pro Cys Ser Ala Val Thr Ile Pro Ala Cys Ala				
		420		425	430
20	Phe Val Asp Leu Asp Ala Glu Leu Cys Leu Gly Gly Pro Gly Ala Ala				
		435		440	445
	Phe Leu Tyr Leu Val Phe Tyr Gln Cys Arg Asp Gln Glu Leu Cys Cys				
		450		455	460
	Val Tyr Val Val Lys Ser Gln Leu Pro Pro Arg Gly Leu Glu Ala Ala				
25	465		470		475
	Leu Glu Arg Leu Phe Gly Arg Leu Arg Ile Thr Asn Thr Ile His Gly				
		485		490	495
	Ala Glu Asp Met Thr Pro Pro Pro Pro Asn Arg Asn Val Asp Phe Pro				
		500		505	510
30	Leu Ala Val Leu Ala Ala Ser Ser Gln Ser Pro Arg Cys Ser Ala Ser				
		515		520	525
	Gln Val Thr Asn Pro Gln Phe Val Asp Arg Leu Tyr Arg Trp Gln Pro				
		530		535	540
	Asp Leu Arg Gly Arg Pro Thr Ala Arg Thr Cys Thr Tyr Ala Ala Phe				
35	545		550		555
	Ala Glu Leu Gly Val Met Pro Asp Asn Ser Pro Arg Cys Leu His Arg				
		565		570	575
	Thr Glu Arg Phe Gly Ala Val Gly Val Pro Val Val Ile Gly Val Val				
		580		585	590
40	Trp Arg Pro Gly Gly Trp Arg Ala Cys Ala				
		595		600	

(2) INFORMATION FOR SEQ ID NO:277:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 515 amino acids

(B) TYPE: amino acid

5 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

Met His Arg Pro Ser Gln Ser Asp Thr Gly Thr His Gln Arg Arg Lys  
 1 5 10 15  
 Pro Pro Ser Pro Val Arg Val Arg Val Phe Ser Gly Gly Val Phe Phe  
 15 20 25 30  
 Leu Ser Ala Leu Leu Pro Pro His Leu His His Pro Pro Pro Thr Trp  
 35 40 45  
 Leu Ala Ile Gly Gly Lys Thr Met Lys Thr Lys Pro Leu Pro Thr Ala  
 50 55 60  
 20 Pro Met Ala Trp Ala Glu Ser Ala Val Glu Thr Thr Thr Ser Pro Arg  
 65 70 75 80  
 Glu Leu Ala Gly His Ala Pro Leu Arg Arg Val Leu Arg Pro Pro Ile  
 85 90 95  
 Ala Arg Arg Asp Gly Pro Val Leu Leu Gly Asp Arg Ala Pro Arg Arg  
 25 100 105 110  
 Thr Ala Ser Thr Met Trp Leu Leu Gly Ile Asp Pro Ala Glu Ser Ser  
 115 120 125  
 Pro Gly Thr Arg Ala Thr Arg Asp Asp Thr Glu Gln Ala Val Asp Lys  
 130 135 140  
 30 Ile Leu Arg Gly Ala Arg Arg Ala Gly Gly Leu Thr Val Pro Gly Ala  
 145 150 155 160  
 Pro Arg Tyr His Leu Thr Arg Gln Val Thr Leu Thr Asp Leu Cys Gln  
 165 170 175  
 Pro Asn Ala Glu Arg Ala Gly Ala Leu Leu Leu Ala Leu Arg His Pro  
 35 180 185 190  
 Thr Asp Leu Pro His Leu Ala Arg His Arg Ala Pro Pro Gly Arg Gln  
 195 200 205  
 Thr Glu Arg Leu Ala Glu Ala Trp Gly Gln Leu Leu Glu Ala Ser Ala  
 210 215 220  
 40 Leu Gly Ser Gly Arg Ala Glu Ser Gly Cys Ala Arg Ala Gly Leu Val  
 225 230 235 240  
 Ser Phe Asn Phe Leu Val Ala Ala Cys Ala Ala Ala Tyr Asp Ala Arg  
 245 250 255

```

Asp Ala Ala Glu Ala Val Arg Ala His Ile Thr Thr Asn Tyr Gly Gly
      260                265                270
Thr Arg Ala Gly Ala Arg Leu Asp Arg Phe Ser Glu Cys Leu Arg Ala
      275                280                285
5 Met Val His Thr His Val Phe Phe Val Met Arg Phe Phe Gly Gly Leu
      290                295                300
Val Ser Trp Val Thr Gln Asp Glu Leu Ala Ser Val Thr Ala Val Cys
      305                310                315                320
Ser Gly Pro Gln Glu Ala Thr His Thr Gly His Pro Gly Arg Pro Cys
10      325                330                335
Ser Ala Val Thr Ile Pro Ala Cys Ala Phe Val Asp Leu Asp Ala Glu
      340                345                350
Leu Cys Leu Gly Gly Pro Gly Ala Ala Phe Leu Tyr Leu Val Phe Tyr
      355                360                365
15 Gln Cys Arg Asp Gln Glu Leu Cys Cys Val Tyr Val Val Lys Ser Gln
      370                375                380
Leu Pro Pro Arg Gly Leu Glu Ala Ala Leu Glu Arg Leu Phe Gly Arg
      385                390                395                400
Leu Arg Ile Thr Asn Thr Ile His Gly Ala Glu Asp Met Thr Pro Pro
20      405                410                415
Pro Pro Asn Arg Asn Val Asp Phe Pro Leu Ala Val Leu Ala Ala Ser
      420                425                430
Ser Gln Ser Pro Arg Cys Ser Ala Ser Gln Val Thr Asn Pro Gln Phe
      435                440                445
25 Val Asp Arg Leu Tyr Arg Trp Gln Pro Asp Leu Arg Gly Arg Pro Thr
      450                455                460
Ala Arg Thr Cys Thr Tyr Ala Ala Phe Ala Glu Leu Gly Val Met Pro
      465                470                475                480
Asp Asn Ser Pro Arg Cys Leu His Arg Thr Glu Arg Phe Gly Ala Val
30      485                490                495
Gly Val Pro Val Val Ile Gly Val Val Trp Arg Pro Gly Gly Trp Arg
      500                505                510
Ala Cys Ala
      515
35

```

(2) INFORMATION FOR SEQ ID NO:278:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 460 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:

```

5  Met Lys Thr Lys Pro Leu Pro Thr Ala Pro Met Ala Trp Ala Glu Ser
   1           5           10           15
Ala Val Glu Thr Thr Thr Ser Pro Arg Glu Leu Ala Gly His Ala Pro
   20           25           30
10 Leu Arg Arg Val Leu Arg Pro Pro Ile Ala Arg Arg Asp Gly Pro Val
   35           40           45
Leu Leu Gly Asp Arg Ala Pro Arg Arg Thr Ala Ser Thr Met Trp Leu
   50           55           60
Leu Gly Ile Asp Pro Ala Glu Ser Ser Pro Gly Thr Arg Ala Thr Arg
   65           70           75           80
15 Asp Asp Thr Glu Gln Ala Val Asp Lys Ile Leu Arg Gly Ala Arg Arg
   85           90           95
Ala Gly Gly Leu Thr Val Pro Gly Ala Pro Arg Tyr His Leu Thr Arg
   100          105          110
Gln Val Thr Leu Thr Asp Leu Cys Gln Pro Asn Ala Glu Arg Ala Gly
   115          120          125
20 Ala Leu Leu Leu Ala Leu Arg His Pro Thr Asp Leu Pro His Leu Ala
   130          135          140
Arg His Arg Ala Pro Pro Gly Arg Gln Thr Glu Arg Leu Ala Glu Ala
   145          150          155          160
25 Trp Gly Gln Leu Leu Glu Ala Ser Ala Leu Gly Ser Gly Arg Ala Glu
   165          170          175
Ser Gly Cys Ala Arg Ala Gly Leu Val Ser Phe Asn Phe Leu Val Ala
   180          185          190
Ala Cys Ala Ala Ala Tyr Asp Ala Arg Asp Ala Ala Glu Ala Val Arg
   195          200          205
30 Ala His Ile Thr Thr Asn Tyr Gly Gly Thr Arg Ala Gly Ala Arg Leu
   210          215          220
Asp Arg Phe Ser Glu Cys Leu Arg Ala Met Val His Thr His Val Phe
   225          230          235          240
35 Phe Val Met Arg Phe Phe Gly Gly Leu Val Ser Trp Val Thr Gln Asp
   245          250          255
Glu Leu Ala Ser Val Thr Ala Val Cys Ser Gly Pro Gln Glu Ala Thr
   260          265          270
His Thr Gly His Pro Gly Arg Pro Cys Ser Ala Val Thr Ile Pro Ala
   275          280          285
40 Cys Ala Phe Val Asp Leu Asp Ala Glu Leu Cys Leu Gly Gly Pro Gly
   290          295          300
Ala Ala Phe Leu Tyr Leu Val Phe Tyr Gln Cys Arg Asp Gln Glu Leu

```

305                      310                      315                      320  
 Cys Cys Val Tyr Val Val Lys Ser Gln Leu Pro Pro Arg Gly Leu Glu  
                          325                      330                      335  
 Ala Ala Leu Glu Arg Leu Phe Gly Arg Leu Arg Ile Thr Asn Thr Ile  
 5                      340                      345                      350  
 His Gly Ala Glu Asp Met Thr Pro Pro Pro Pro Asn Arg Asn Val Asp  
                          355                      360                      365  
 Phe Pro Leu Ala Val Leu Ala Ala Ser Ser Gln Ser Pro Arg Cys Ser  
                          370                      375                      380  
 10    Ala Ser Gln Val Thr Asn Pro Gln Phe Val Asp Arg Leu Tyr Arg Trp  
                          385                      390                      395                      400  
 Gln Pro Asp Leu Arg Gly Arg Pro Thr Ala Arg Thr Cys Thr Tyr Ala  
                          405                      410                      415  
 Ala Phe Ala Glu Leu Gly Val Met Pro Asp Asn Ser Pro Arg Cys Leu  
 15                      420                      425                      430  
 His Arg Thr Glu Arg Phe Gly Ala Val Gly Val Pro Val Val Ile Gly  
                          435                      440                      445  
 Val Val Trp Arg Pro Gly Gly Trp Arg Ala Cys Ala  
                          450                      455                      460

20

(2) INFORMATION FOR SEQ ID NO:279:

(i) SEQUENCE CHARACTERISTICS:

25

- (A) LENGTH: 452 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:279:

Met Ala Val Val Cys Gly Ser Gly Leu Arg Leu Arg Pro Phe His Pro  
   1                      5                      10                      15  
 35    Pro Ser Pro Ser Phe Phe Val Leu Arg Ala Leu Ile Arg Ala Gly Pro  
                          20                      25                      30  
 Gly Pro Phe Ala Asp Arg Ala Pro Ser Gly Pro Gly Cys Gly Met Cys  
                          35                      40                      45  
 Arg Gly Asp Ser Pro Gly Val Ala Gly Gly Ser Gly Glu His Cys Leu  
 40                      50                      55                      60  
 Gly Gly Asp Asp Gly Asp Asp Gly Arg Pro Arg Leu Ala Cys Val Gly  
   65                      70                      75                      80  
 Ala Ile Arg Phe Ala His Leu Trp Leu Gln Ala Thr Thr Leu Gly Phe

		85		90		95
	Val Gly Ser Val Val Leu Ser Arg Gly Pro Tyr Ala Asp Ala Met Ser					
	100		105		110	
5	Gly Ala Phe Val Ile Gly Ser Thr Gly Leu Gly Phe Leu Arg Ala Pro					
	115		120		125	
	Pro Ala Phe Ala Arg Pro Pro Thr Arg Val Cys Ala Trp Leu Arg Leu					
	130		135		140	
	Val Gly Gly Gly Ala Ala Val Trp Ser Leu Gly Glu Ala Gly Ala Pro					
	145		150		155	160
10	Pro Gly Val Pro Gly Pro Ala Thr Gln Cys Leu Ala Leu Gly Ala Ala					
	165		170		175	
	Tyr Ala Ala Leu Leu Val Leu Ala Asp Asp Val His Pro Leu Phe Leu					
	180		185		190	
	Leu Ala Pro Arg Pro Leu Phe Val Gly Thr Leu Gly Val Val Val Gly					
15	195		200		205	
	Gly Leu Thr Ile Gly Gly Ser Ala Arg Tyr Trp Trp Ile Asp Pro Arg					
	210		215		220	
	Ala Ala Ala Ala Leu Thr Ala Ala Val Val Ala Gly Leu Gly Thr Thr					
	225		230		235	240
20	Ala Ala Gly Asp Ser Phe Ser Lys Ala Cys Pro Arg His Arg Arg Phe					
	245		250		255	
	Cys Val Val Ser Ala Val Glu Ser Pro Pro Pro Arg Tyr Ala Pro Glu					
	260		265		270	
	Asp Ala Glu Arg Pro Thr Asp His Gly Pro Leu Leu Pro Ser Thr His					
25	275		280		285	
	His Gln Arg Ser Pro Arg Val Cys Gly Asp Gly Ala Ala Arg Pro Glu					
	290		295		300	
	Asn Ile Trp Val Pro Val Val Thr Phe Ala Gly Ala Leu Ala Ala Cys					
	305		310		315	320
30	Ala Arg Ser Asp Ala Ala Pro Ser Gly Pro Val Leu Pro Leu Trp Pro					
	325		330		335	
	Gln Val Phe Val Gly Gly His Ala Ala Ala Gly Leu Thr Glu Leu Cys					
	340		345		350	
	Gln Thr Leu Ala Pro Arg Asp Leu Thr Asp Pro Leu Leu Phe Ala Tyr					
35	355		360		365	
	Val Gly Phe Gln Val Val Asn His Gly Leu Met Phe Val Val Pro Asp					
	370		375		380	
	Ile Ala Val Tyr Ala Met Leu Gly Gly Ala Val Trp Ile Ser Leu Thr					
	385		390		395	400
40	Gln Val Leu Gly Leu Arg Arg Arg Leu His Lys Asp Pro Asp Ala Gly					
	405		410		415	
	Pro Trp Ala Ala Ala Thr Leu Arg Gly Leu Phe Phe Ser Val Tyr Ala					
	420		425		430	

Leu Gly Phe Ala Ala Gly Val Leu Val Arg Pro Arg Met Ala Ala Ser  
 435 440 445  
 Arg Arg Ser Gly  
 450

5

## (2) INFORMATION FOR SEQ ID NO:280:

## (i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 406 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

Met Cys Arg Gly Asp Ser Pro Gly Val Ala Gly Gly Ser Gly Glu His  
 1 5 10 15  
 20 Cys Leu Gly Gly Asp Asp Gly Asp Asp Gly Arg Pro Arg Leu Ala Cys  
 20 25 30  
 Val Gly Ala Ile Arg Phe Ala His Leu Trp Leu Gln Ala Thr Thr Leu  
 35 40 45  
 Gly Phe Val Gly Ser Val Val Leu Ser Arg Gly Pro Tyr Ala Asp Ala  
 25 50 55 60  
 Met Ser Gly Ala Phe Val Ile Gly Ser Thr Gly Leu Gly Phe Leu Arg  
 65 70 75 80  
 Ala Pro Pro Ala Phe Ala Arg Pro Pro Thr Arg Val Cys Ala Trp Leu  
 85 90 95  
 30 Arg Leu Val Gly Gly Gly Ala Ala Val Trp Ser Leu Gly Glu Ala Gly  
 100 105 110  
 Ala Pro Pro Gly Val Pro Gly Pro Ala Thr Gln Cys Leu Ala Leu Gly  
 115 120 125  
 Ala Ala Tyr Ala Ala Leu Leu Val Leu Ala Asp Asp Val His Pro Leu  
 35 130 135 140  
 Phe Leu Leu Ala Pro Arg Pro Leu Phe Val Gly Thr Leu Gly Val Val  
 145 150 155 160  
 Val Gly Gly Leu Thr Ile Gly Gly Ser Ala Arg Tyr Trp Trp Ile Asp  
 165 170 175  
 40 Pro Arg Ala Ala Ala Ala Leu Thr Ala Ala Val Val Ala Gly Leu Gly  
 180 185 190  
 Thr Thr Ala Ala Gly Asp Ser Phe Ser Lys Ala Cys Pro Arg His Arg  
 195 200 205

688

Arg Phe Cys Val Val Ser Ala Val Glu Ser Pro Pro Pro Arg Tyr Ala  
 210 215 220  
 Pro Glu Asp Ala Glu Arg Pro Thr Asp His Gly Pro Leu Leu Pro Ser  
 225 230 235 240  
 5 Thr His His Gln Arg Ser Pro Arg Val Cys Gly Asp Gly Ala Ala Arg  
 245 250 255  
 Pro Glu Asn Ile Trp Val Pro Val Val Thr Phe Ala Gly Ala Leu Ala  
 260 265 270  
 10 Ala Cys Ala Arg Ser Asp Ala Ala Pro Ser Gly Pro Val Leu Pro Leu  
 275 280 285  
 Trp Pro Gln Val Phe Val Gly Gly His Ala Ala Ala Gly Leu Thr Glu  
 290 295 300  
 Leu Cys Gln Thr Leu Ala Pro Arg Asp Leu Thr Asp Pro Leu Leu Phe  
 305 310 315 320  
 15 Ala Tyr Val Gly Phe Gln Val Val Asn His Gly Leu Met Phe Val Val  
 325 330 335  
 Pro Asp Ile Ala Val Tyr Ala Met Leu Gly Gly Ala Val Trp Ile Ser  
 340 345 350  
 Leu Thr Gln Val Leu Gly Leu Arg Arg Arg Leu His Lys Asp Pro Asp  
 20 355 360 365  
 Ala Gly Pro Trp Ala Ala Ala Thr Leu Arg Gly Leu Phe Phe Ser Val  
 370 375 380  
 Tyr Ala Leu Gly Phe Ala Ala Gly Val Leu Val Arg Pro Arg Met Ala  
 385 390 395 400  
 25 Ala Ser Arg Arg Ser Gly  
 405

## (2) INFORMATION FOR SEQ ID NO:281:

## 30 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 644 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

35

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:281:

40 Met Gly Thr Glu Asp Cys Asp His Glu Gly Arg Ser Val Ala Ala Pro  
 1 5 10 15  
 Val Glu Val Met Ala Leu Tyr Ala Thr Asp Gly Cys Val Ile Thr Ser  
 20 25 30  
 689



Ser Leu Ala Leu Leu Thr Asn Cys Leu Leu Gly Ala Glu Pro Leu Tyr  
 35 40 45  
 Ile Phe Ser Tyr Asp Ala Tyr Arg Pro Asp Ala Pro Asn Gly Pro Thr  
 50 55 60  
 5 Gly Ala Pro Thr Glu Gln Glu Arg Phe Glu Gly Ser Arg Ala Leu Tyr  
 65 70 75 80  
 Arg Asp Ala Gly Gln Gly Asp Ser Phe Arg Val Thr Phe Cys Leu Leu  
 85 90 95  
 Gly Thr Glu Val Gly Val Thr His His Pro Lys Gly Arg Trp Met Phe  
 10 100 105 110  
 Val Cys Arg Phe Glu Arg Ala Asp Asp Val Ala Val Leu Gln Asp Ala  
 115 120 125  
 Leu Gly Arg Gly Thr Pro Leu Leu Pro Ala His Ile Thr Ala Thr Leu  
 130 135 140  
 15 Asp Leu Glu Ala Thr Phe Ala Leu His Ala Asn Ile Ile Met Ala Leu  
 145 150 155 160  
 Thr Val Ala Ile Val His Asn Ala Pro Ala Arg Ile Gly Ser Gly Ser  
 165 170 175  
 Thr Ala Pro Leu Tyr Glu Pro Gly Glu Ser Met Arg Ser Val Val Gly  
 180 185 190  
 20 Arg Met Ser Leu Gly Gln Arg Gly Leu Thr Thr Leu Phe Val His His  
 195 200 205  
 Glu Ala Arg Val Leu Ala Ala Tyr Arg Arg Ala Tyr Tyr Gly Ser Ala  
 210 215 220  
 25 Gln Ser Pro Phe Trp Phe Leu Ser Lys Phe Gly Pro Asp Glu Lys Ser  
 225 230 235 240  
 Leu Val Leu Ala Ala Arg Tyr Tyr Val Leu Gln Ala Pro Arg Leu Gly  
 245 250 255  
 Gly Ala Gly Ala Thr Tyr Asp Leu Gln Ala Val Lys Asp Ile Cys Ala  
 260 265 270  
 30 Thr Tyr Ala Ile Pro His Asp Pro Arg Pro Asp Thr Leu Ser Ala Ala  
 275 280 285  
 Ser Leu Thr Ser Phe Ala Ala Ile Thr Arg Phe Cys Cys Thr Ser Gln  
 290 295 300  
 35 Tyr Ser Arg Gly Ala Ala Ala Ala Gly Phe Pro Leu Tyr Val Glu Arg  
 305 310 315 320  
 Arg Ile Ala Ala Asp Val Arg Glu Thr Gly Ala Leu Glu Lys Phe Ile  
 325 330 335  
 Ala His Asp Arg Ser Cys Leu Arg Val Ser Asp Arg Glu Phe Ile Thr  
 340 345 350  
 40 Tyr Ile Tyr Leu Ala His Phe Glu Cys Phe Ser Pro Pro Arg Leu Ala  
 355 360 365  
 Thr His Leu Arg Ala Val Thr Thr His Asp Pro Ser Pro Ala Ala Ser  
 690

370 375 380  
 Thr Glu Gln Pro Ser Pro Leu Gly Arg Glu Ala Val Glu Gln Phe Phe  
 385 390 395 400  
 Arg His Val Arg Ala Gln Leu Asn Ile Arg Glu Tyr Val Lys Gln Asn  
 5 405 410 415  
 Val Thr Pro Arg Glu Thr Ala Gly Asp Ala Ala Ala Tyr Leu Arg  
 420 425 430  
 Ala Arg Thr Tyr Ala Pro Ala Ala Leu Thr Pro Ala Pro Ala Tyr Cys  
 435 440 445  
 10 Gly Val Ala Asp Ser Ser Thr Lys Met Met Gly Arg Leu Ala Glu Ala  
 450 455 460  
 Glu Arg Leu Leu Val Pro His Gly Trp Pro Ala Phe Ala Pro Thr Thr  
 465 470 475 480  
 Pro Gly Asp Asp Ala Gly Gly Gly Thr Ala Ala Pro Gln Thr Cys Gly  
 15 485 490 495  
 Ile Val Lys Arg Leu Leu Lys Leu Ala Ala Thr Glu Gln Gln Gly Thr  
 500 505 510  
 Thr Pro Pro Ala Ile Ala Ala Leu Met Gln Asp Ala Ser Val Gln Thr  
 515 520 525  
 20 Pro Leu Pro Val Tyr Arg Ile Thr Met Ser Pro Thr Gly Gln Ala Phe  
 530 535 540  
 Ala Ala Ala Ala Arg Asp Asp Trp Ala Arg Val Thr Arg Asp Ala Arg  
 545 550 555 560  
 Pro Pro Glu Ala Thr Val Val Ala Asp Ala Ala Ala Pro Glu Pro  
 25 565 570 575  
 Gly Ala Leu Gly Arg Arg Leu Thr Arg Arg Ile Cys Arg Pro Ala Pro  
 580 585 590  
 Pro Pro Gly Arg Pro Gly Arg Arg Gly Pro Asp Val Arg Glu Pro Gln  
 595 600 605  
 30 Arg Asp Leu Gln Arg Arg Ala Gly Arg Tyr Glu His His Pro Gly Ser  
 610 615 620  
 Gly His Arg Pro Glu Gly Ala Arg Pro Leu Ser Pro Ala Pro Arg Gly  
 625 630 635 640  
 Pro Gly Ser Leu  
 35

## (2) INFORMATION FOR SEQ ID NO:282:

## (i) SEQUENCE CHARACTERISTICS:

- 40
- (A) LENGTH: 715 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:282:

5  
 Met Gly Ala Gly Lys Ser Ala Leu Thr Thr Ala Arg Ala Ser Cys Ser  
 1 5 10 15  
 Arg Gly Ser Xaa Ser Glu Gly Gly Ala Ala Ala Arg Ile Ile Ser Tyr  
 20 25 30  
 10 Cys Cys Ser Ser Gly Arg Val Pro Gln Pro His Ser Thr Pro Ser Arg  
 35 40 45  
 Asp Ala Ile Pro Glu His Arg Ser Ala Pro Ala Phe Pro His Pro Thr  
 50 55 60  
 Pro Ser Gly Phe Ala Gly Ala Met Gly Thr Glu Asp Cys Asp His Glu  
 15 65 70 75 80  
 Gly Arg Ser Val Ala Ala Pro Val Glu Val Met Ala Leu Tyr Ala Thr  
 85 90 95  
 Asp Gly Cys Val Ile Thr Ser Ser Leu Ala Leu Leu Thr Asn Cys Leu  
 100 105 110  
 20 Leu Gly Ala Glu Pro Leu Tyr Ile Phe Ser Tyr Asp Ala Tyr Arg Pro  
 115 120 125  
 Asp Ala Pro Asn Gly Pro Thr Gly Ala Pro Thr Glu Gln Glu Arg Phe  
 130 135 140  
 Glu Gly Ser Arg Ala Leu Tyr Arg Asp Ala Gly Gln Gly Asp Ser Phe  
 25 145 150 155 160  
 Arg Val Thr Phe Cys Leu Leu Gly Thr Glu Val Gly Val Thr His His  
 165 170 175  
 Pro Lys Gly Arg Trp Met Phe Val Cys Arg Phe Glu Arg Ala Asp Asp  
 180 185 190  
 30 Val Ala Val Leu Gln Asp Ala Leu Gly Arg Gly Thr Pro Leu Leu Pro  
 195 200 205  
 Ala His Ile Thr Ala Thr Leu Asp Leu Glu Ala Thr Phe Ala Leu His  
 210 215 220  
 Ala Asn Ile Ile Met Ala Leu Thr Val Ala Ile Val His Asn Ala Pro  
 35 225 230 235 240  
 Ala Arg Ile Gly Ser Gly Ser Thr Ala Pro Leu Tyr Glu Pro Gly Glu  
 245 250 255  
 Ser Met Arg Ser Val Val Gly Arg Met Ser Leu Gly Gln Arg Gly Leu  
 260 265 270  
 40 Thr Thr Leu Phe Val His His Glu Ala Arg Val Leu Ala Ala Tyr Arg  
 275 280 285  
 Arg Ala Tyr Tyr Gly Ser Ala Gln Ser Pro Phe Trp Phe Leu Ser Lys  
 290 295 300

Phe Gly Pro Asp Glu Lys Ser Leu Val Leu Ala Ala Arg Tyr Tyr Val  
 305 310 315 320  
 Leu Gln Ala Pro Arg Leu Gly Gly Ala Gly Ala Thr Tyr Asp Leu Gln  
 325 330 335  
 5 Ala Val Lys Asp Ile Cys Ala Thr Tyr Ala Ile Pro His Asp Pro Arg  
 340 345 350  
 Pro Asp Thr Leu Ser Ala Ala Ser Leu Thr Ser Phe Ala Ala Ile Thr  
 355 360 365  
 Arg Phe Cys Cys Thr Ser Gln Tyr Ser Arg Gly Ala Ala Ala Ala Gly  
 10 370 375 380  
 Phe Pro Leu Tyr Val Glu Arg Arg Ile Ala Ala Asp Val Arg Glu Thr  
 385 390 395 400  
 Gly Ala Leu Glu Lys Phe Ile Ala His Asp Arg Ser Cys Leu Arg Val  
 405 410 415  
 15 Ser Asp Arg Glu Phe Ile Thr Tyr Ile Tyr Leu Ala His Phe Glu Cys  
 420 425 430  
 Phe Ser Pro Pro Arg Leu Ala Thr His Leu Arg Ala Val Thr Thr His  
 435 440 445  
 Asp Pro Ser Pro Ala Ala Ser Thr Glu Gln Pro Ser Pro Leu Gly Arg  
 20 450 455 460  
 Glu Ala Val Glu Gln Phe Phe Arg His Val Arg Ala Gln Leu Asn Ile  
 465 470 475 480  
 Arg Glu Tyr Val Lys Gln Asn Val Thr Pro Arg Glu Thr Ala Gly Asp  
 485 490 495  
 25 Ala Ala Ala Ala Tyr Leu Arg Ala Arg Thr Tyr Ala Pro Ala Ala Leu  
 500 505 510  
 Thr Pro Ala Pro Ala Tyr Cys Gly Val Ala Asp Ser Ser Thr Lys Met  
 515 520 525  
 Met Gly Arg Leu Ala Glu Ala Glu Arg Leu Leu Val Pro His Gly Trp  
 30 530 535 540  
 Pro Ala Phe Ala Pro Thr Thr Pro Gly Asp Asp Ala Gly Gly Gly Thr  
 545 550 555 560  
 Ala Ala Pro Gln Thr Cys Gly Ile Val Lys Arg Leu Leu Lys Leu Ala  
 565 570 575  
 35 Ala Thr Glu Gln Gln Gly Thr Thr Pro Pro Ala Ile Ala Ala Leu Met  
 580 585 590  
 Gln Asp Ala Ser Val Gln Thr Pro Leu Pro Val Tyr Arg Ile Thr Met  
 595 600 605  
 Ser Pro Thr Gly Gln Ala Phe Ala Ala Ala Ala Arg Asp Asp Trp Ala  
 40 610 615 620  
 Arg Val Thr Arg Asp Ala Arg Pro Pro Glu Ala Thr Val Val Ala Asp  
 625 630 635 640  
 Ala Ala Ala Ala Pro Glu Pro Gly Ala Leu Gly Arg Arg Leu Thr Arg

645 650 655  
 Arg Ile Cys Arg Pro Ala Pro Pro Pro Gly Arg Pro Gly Arg Arg Gly  
 660 665 670  
 Pro Asp Val Arg Glu Pro Gln Arg Asp Leu Gln Arg Arg Ala Gly Arg  
 5 675 680 685  
 Tyr Glu His His Pro Gly Ser Gly His Arg Pro Glu Gly Ala Arg Pro  
 690 695 700  
 Leu Ser Pro Ala Pro Arg Gly Pro Gly Ser Leu  
 705 710 715

10

## (2) INFORMATION FOR SEQ ID NO:283:

## (i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 744 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:283:

Met Gln Ala Trp Tyr Val Arg Ala Arg Ala Arg Ala Phe Thr Arg Arg  
 1 5 10 15  
 25 Arg Val Ser Ser Ser Asp Ser Arg Ala Ser Ser Ser Val Met Gly Ala  
 20 25 30  
 Gly Lys Ser Ala Leu Thr Thr Ala Arg Ala Ser Cys Ser Arg Gly Ser  
 35 40 45  
 Xaa Ser Glu Gly Gly Ala Ala Ala Arg Ile Ile Ser Tyr Cys Cys Ser  
 30 50 55 60  
 Ser Gly Arg Val Pro Gln Pro His Ser Thr Pro Ser Arg Asp Ala Ile  
 65 70 75 80  
 Pro Glu His Arg Ser Ala Pro Ala Phe Pro His Pro Thr Pro Ser Gly  
 85 90 95  
 35 Phe Ala Gly Ala Met Gly Thr Glu Asp Cys Asp His Glu Gly Arg Ser  
 100 105 110  
 Val Ala Ala Pro Val Glu Val Met Ala Leu Tyr Ala Thr Asp Gly Cys  
 115 120 125  
 Val Ile Thr Ser Ser Leu Ala Leu Leu Thr Asn Cys Leu Leu Gly Ala  
 40 130 135 140  
 Glu Pro Leu Tyr Ile Phe Ser Tyr Asp Ala Tyr Arg Pro Asp Ala Pro  
 145 150 155 160  
 Asn Gly Pro Thr Gly Ala Pro Thr Glu Gln Glu Arg Phe Glu Gly Ser

[illegible]

Val Lys Gln Asn Val Thr Pro Arg Glu Thr Ala Gly Asp Ala Ala Ala  
 515 520 525  
 Ala Tyr Leu Arg Ala Arg Thr Tyr Ala Pro Ala Ala Leu Thr Pro Ala  
 530 535 540  
 5 Pro Ala Tyr Cys Gly Val Ala Asp Ser Ser Thr Lys Met Met Gly Arg  
 545 550 555 560  
 Leu Ala Glu Ala Glu Arg Leu Leu Val Pro His Gly Trp Pro Ala Phe  
 565 570 575  
 Ala Pro Thr Thr Pro Gly Asp Asp Ala Gly Gly Gly Thr Ala Ala Pro  
 10 580 585 590  
 Gln Thr Cys Gly Ile Val Lys Arg Leu Leu Lys Leu Ala Ala Thr Glu  
 595 600 605  
 Gln Gln Gly Thr Thr Pro Pro Ala Ile Ala Ala Leu Met Gln Asp Ala  
 610 615 620  
 15 Ser Val Gln Thr Pro Leu Pro Val Tyr Arg Ile Thr Met Ser Pro Thr  
 625 630 635 640  
 Gly Gln Ala Phe Ala Ala Ala Ala Arg Asp Asp Trp Ala Arg Val Thr  
 645 650 655  
 Arg Asp Ala Arg Pro Pro Glu Ala Thr Val Val Ala Asp Ala Ala Ala  
 20 660 665 670  
 Ala Pro Glu Pro Gly Ala Leu Gly Arg Arg Leu Thr Arg Arg Ile Cys  
 675 680 685  
 Arg Pro Ala Pro Pro Pro Gly Arg Pro Gly Arg Arg Gly Pro Asp Val  
 690 695 700  
 25 Arg Glu Pro Gln Arg Asp Leu Gln Arg Arg Ala Gly Arg Tyr Glu His  
 705 710 715 720  
 His Pro Gly Ser Gly His Arg Pro Glu Gly Ala Arg Pro Leu Ser Pro  
 725 730 735  
 Ala Pro Arg Gly Pro Gly Ser Leu  
 30 740

## (2) INFORMATION FOR SEQ ID NO:284:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 762 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:284:

	Met	Val	Glu	Pro	Ser	Ser	Pro	Gly	Trp	Trp	Arg	Ala	Ser	Leu	Ser	Arg
	1				5				10					15		
	Leu	Thr	Met	Gln	Ala	Trp	Tyr	Val	Arg	Ala	Arg	Ala	Arg	Ala	Phe	Thr
				20					25					30		
5	Arg	Arg	Arg	Val	Ser	Ser	Ser	Asp	Ser	Arg	Ala	Ser	Ser	Ser	Val	Met
				35				40						45		
	Gly	Ala	Gly	Lys	Ser	Ala	Leu	Thr	Thr	Ala	Arg	Ala	Ser	Cys	Ser	Arg
		50					55					60				
	Gly	Ser	Xaa	Ser	Glu	Gly	Gly	Ala	Ala	Ala	Arg	Ile	Ile	Ser	Tyr	Cys
10	65					70					75				80	
	Cys	Ser	Ser	Gly	Arg	Val	Pro	Gln	Pro	His	Ser	Thr	Pro	Ser	Arg	Asp
					85					90					95	
	Ala	Ile	Pro	Glu	His	Arg	Ser	Ala	Pro	Ala	Phe	Pro	His	Pro	Thr	Pro
				100					105					110		
15	Ser	Gly	Phe	Ala	Gly	Ala	Met	Gly	Thr	Glu	Asp	Cys	Asp	His	Glu	Gly
				115				120					125			
	Arg	Ser	Val	Ala	Ala	Pro	Val	Glu	Val	Met	Ala	Leu	Tyr	Ala	Thr	Asp
		130					135					140				
	Gly	Cys	Val	Ile	Thr	Ser	Ser	Leu	Ala	Leu	Leu	Thr	Asn	Cys	Leu	Leu
20	145					150					155				160	
	Gly	Ala	Glu	Pro	Leu	Tyr	Ile	Phe	Ser	Tyr	Asp	Ala	Tyr	Arg	Pro	Asp
					165					170				175		
	Ala	Pro	Asn	Gly	Pro	Thr	Gly	Ala	Pro	Thr	Glu	Gln	Glu	Arg	Phe	Glu
				180				185					190			
25	Gly	Ser	Arg	Ala	Leu	Tyr	Arg	Asp	Ala	Gly	Gln	Gly	Asp	Ser	Phe	Arg
				195				200					205			
	Val	Thr	Phe	Cys	Leu	Leu	Gly	Thr	Glu	Val	Gly	Val	Thr	His	His	Pro
		210					215					220				
	Lys	Gly	Arg	Trp	Met	Phe	Val	Cys	Arg	Phe	Glu	Arg	Ala	Asp	Asp	Val
30	225					230				235					240	
	Ala	Val	Leu	Gln	Asp	Ala	Leu	Gly	Arg	Gly	Thr	Pro	Leu	Leu	Pro	Ala
					245					250				255		
	His	Ile	Thr	Ala	Thr	Leu	Asp	Leu	Glu	Ala	Thr	Phe	Ala	Leu	His	Ala
				260				265					270			
35	Asn	Ile	Ile	Met	Ala	Leu	Thr	Val	Ala	Ile	Val	His	Asn	Ala	Pro	Ala
				275				280					285			
	Arg	Ile	Gly	Ser	Gly	Ser	Thr	Ala	Pro	Leu	Tyr	Glu	Pro	Gly	Glu	Ser
		290					295					300				
	Met	Arg	Ser	Val	Val	Gly	Arg	Met	Ser	Leu	Gly	Gln	Arg	Gly	Leu	Thr
40	305					310					315				320	
	Thr	Leu	Phe	Val	His	His	Glu	Ala	Arg	Val	Leu	Ala	Ala	Tyr	Arg	Arg
					325					330				335		
	Ala	Tyr	Tyr	Gly	Ser	Ala	Gln	Ser	Pro	Phe	Trp	Phe	Leu	Ser	Lys	Phe



		340		345		350	
	Gly	Pro	Asp	Glu	Lys	Ser	Leu
		355		360		365	
	Gln	Ala	Pro	Arg	Leu	Gly	Gly
5		370		375		380	
	Val	Lys	Asp	Ile	Cys	Ala	Thr
	385			390		395	
	Asp	Thr	Leu	Ser	Ala	Ser	Leu
		405		410		415	
10	Phe	Cys	Cys	Thr	Ser	Gln	Tyr
		420		425		430	
	Pro	Leu	Tyr	Val	Glu	Arg	Arg
		435		440		445	
	Ala	Leu	Glu	Lys	Phe	Ile	Ala
15		450		455		460	
	Asp	Arg	Glu	Phe	Ile	Thr	Tyr
	465			470		475	
	Ser	Pro	Pro	Arg	Leu	Ala	Thr
		485		490		495	
20	Pro	Ser	Pro	Ala	Ala	Ser	Thr
		500		505		510	
	Ala	Val	Glu	Gln	Phe	Phe	Arg
		515		520		525	
	Glu	Tyr	Val	Lys	Gln	Asn	Val
25		530		535		540	
	Ala	Ala	Ala	Tyr	Leu	Arg	Ala
	545			550		555	
	Pro	Ala	Pro	Ala	Tyr	Cys	Gly
		565		570		575	
30	Gly	Arg	Leu	Ala	Glu	Ala	Glu
		580		585		590	
	Ala	Phe	Ala	Pro	Thr	Thr	Pro
		595		600		605	
	Ala	Pro	Gln	Thr	Cys	Gly	Ile
35		610		615		620	
	Thr	Glu	Gln	Gln	Gly	Thr	Thr
	625			630		635	
	Asp	Ala	Ser	Val	Gln	Thr	Pro
		645		650		655	
40	Pro	Thr	Gly	Gln	Ala	Phe	Ala
		660		665		670	
	Val	Thr	Arg	Asp	Ala	Arg	Pro
		675		680		685	

Ala Ala Ala Pro Glu Pro Gly Ala Leu Gly Arg Arg Leu Thr Arg Arg  
 690 695 700  
 Ile Cys Arg Pro Ala Pro Pro Pro Gly Arg Pro Gly Arg Arg Gly Pro  
 705 710 715 720  
 5 Asp Val Arg Glu Pro Gln Arg Asp Leu Gln Arg Arg Ala Gly Arg Tyr  
 725 730 735  
 Glu His His Pro Gly Ser Gly His Arg Pro Glu Gly Ala Arg Pro Leu  
 740 745 750  
 Ser Pro Ala Pro Arg Gly Pro Gly Ser Leu  
 10 755 760

## (2) INFORMATION FOR SEQ ID NO:285:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 781 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:285:

Met His Val Ser Ala Arg Arg Arg Ile Leu Ser Arg Cys Ala Ala Thr  
 25 1 5 10 15  
 Ala Pro Ser Met Val Glu Pro Ser Ser Pro Gly Trp Trp Arg Ala Ser  
 20 25 30  
 Leu Ser Arg Leu Thr Met Gln Ala Trp Tyr Val Arg Ala Arg Ala Arg  
 35 40 45  
 30 Ala Phe Thr Arg Arg Arg Val Ser Ser Ser Asp Ser Arg Ala Ser Ser  
 50 55 60  
 Ser Val Met Gly Ala Gly Lys Ser Ala Leu Thr Thr Ala Arg Ala Ser  
 65 70 75 80  
 Cys Ser Arg Gly Ser Xaa Ser Glu Gly Gly Ala Ala Ala Arg Ile Ile  
 35 85 90 95  
 Ser Tyr Cys Cys Ser Ser Gly Arg Val Pro Gln Pro His Ser Thr Pro  
 100 105 110  
 Ser Arg Asp Ala Ile Pro Glu His Arg Ser Ala Pro Ala Phe Pro His  
 115 120 125  
 40 Pro Thr Pro Ser Gly Phe Ala Gly Ala Met Gly Thr Glu Asp Cys Asp  
 130 135 140  
 His Glu Gly Arg Ser Val Ala Ala Pro Val Glu Val Met Ala Leu Tyr  
 145 150 155 160

	Ala Thr Asp Gly Cys Val Ile Thr Ser Ser Leu Ala Leu Leu Thr Asn
	165 170 175
	Cys Leu Leu Gly Ala Glu Pro Leu Tyr Ile Phe Ser Tyr Asp Ala Tyr
	180 185 190
5	Arg Pro Asp Ala Pro Asn Gly Pro Thr Gly Ala Pro Thr Glu Gln Glu
	195 200 205
	Arg Phe Glu Gly Ser Arg Ala Leu Tyr Arg Asp Ala Gly Gln Gly Asp
	210 215 220
10	Ser Phe Arg Val Thr Phe Cys Leu Leu Gly Thr Glu Val Gly Val Thr
	225 230 235 240
	His His Pro Lys Gly Arg Trp Met Phe Val Cys Arg Phe Glu Arg Ala
	245 250 255
	Asp Asp Val Ala Val Leu Gln Asp Ala Leu Gly Arg Gly Thr Pro Leu
	260 265 270
15	Leu Pro Ala His Ile Thr Ala Thr Leu Asp Leu Glu Ala Thr Phe Ala
	275 280 285
	Leu His Ala Asn Ile Ile Met Ala Leu Thr Val Ala Ile Val His Asn
	290 295 300
20	Ala Pro Ala Arg Ile Gly Ser Gly Ser Thr Ala Pro Leu Tyr Glu Pro
	305 310 315 320
	Gly Glu Ser Met Arg Ser Val Val Gly Arg Met Ser Leu Gly Gln Arg
	325 330 335
	Gly Leu Thr Thr Leu Phe Val His His Glu Ala Arg Val Leu Ala Ala
	340 345 350
25	Tyr Arg Arg Ala Tyr Tyr Gly Ser Ala Gln Ser Pro Phe Trp Phe Leu
	355 360 365
	Ser Lys Phe Gly Pro Asp Glu Lys Ser Leu Val Leu Ala Ala Arg Tyr
	370 375 380
30	Tyr Val Leu Gln Ala Pro Arg Leu Gly Gly Ala Gly Ala Thr Tyr Asp
	385 390 395 400
	Leu Gln Ala Val Lys Asp Ile Cys Ala Thr Tyr Ala Ile Pro His Asp
	405 410 415
	Pro Arg Pro Asp Thr Leu Ser Ala Ala Ser Leu Thr Ser Phe Ala Ala
	420 425 430
35	Ile Thr Arg Phe Cys Cys Thr Ser Gln Tyr Ser Arg Gly Ala Ala Ala
	435 440 445
	Ala Gly Phe Pro Leu Tyr Val Glu Arg Arg Ile Ala Ala Asp Val Arg
	450 455 460
40	Glu Thr Gly Ala Leu Glu Lys Phe Ile Ala His Asp Arg Ser Cys Leu
	465 470 475 480
	Arg Val Ser Asp Arg Glu Phe Ile Thr Tyr Ile Tyr Leu Ala His Phe
	485 490 495
	Glu Cys Phe Ser Pro Pro Arg Leu Ala Thr His Leu Arg Ala Val Thr

	500	505	510
	Thr His Asp Pro Ser Pro Ala Ala Ser Thr Glu Gln Pro Ser Pro Leu		
	515	520	525
	Gly Arg Glu Ala Val Glu Gln Phe Phe Arg His Val Arg Ala Gln Leu		
5	530	535	540
	Asn Ile Arg Glu Tyr Val Lys Gln Asn Val Thr Pro Arg Glu Thr Ala		
	545	550	555
	Gly Asp Ala Ala Ala Tyr Leu Arg Ala Arg Thr Tyr Ala Pro Ala		
	565	570	575
10	Ala Leu Thr Pro Ala Pro Ala Tyr Cys Gly Val Ala Asp Ser Ser Thr		
	580	585	590
	Lys Met Met Gly Arg Leu Ala Glu Ala Glu Arg Leu Leu Val Pro His		
	595	600	605
	Gly Trp Pro Ala Phe Ala Pro Thr Thr Pro Gly Asp Asp Ala Gly Gly		
15	610	615	620
	Gly Thr Ala Ala Pro Gln Thr Cys Gly Ile Val Lys Arg Leu Leu Lys		
	625	630	635
	Leu Ala Ala Thr Glu Gln Gln Gly Thr Thr Pro Pro Ala Ile Ala Ala		
	645	650	655
20	Leu Met Gln Asp Ala Ser Val Gln Thr Pro Leu Pro Val Tyr Arg Ile		
	660	665	670
	Thr Met Ser Pro Thr Gly Gln Ala Phe Ala Ala Ala Arg Asp Asp		
	675	680	685
	Trp Ala Arg Val Thr Arg Asp Ala Arg Pro Pro Glu Ala Thr Val Val		
25	690	695	700
	Ala Asp Ala Ala Ala Pro Glu Pro Gly Ala Leu Gly Arg Arg Leu		
	705	710	715
	Thr Arg Arg Ile Cys Arg Pro Ala Pro Pro Pro Gly Arg Pro Gly Arg		
	725	730	735
30	Arg Gly Pro Asp Val Arg Glu Pro Gln Arg Asp Leu Gln Arg Arg Ala		
	740	745	750
	Gly Arg Tyr Glu His His Pro Gly Ser Gly His Arg Pro Glu Gly Ala		
	755	760	765
	Arg Pro Leu Ser Pro Ala Pro Arg Gly Pro Gly Ser Leu		
35	770	775	780

## (2) INFORMATION FOR SEQ ID NO:286:

## (i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 784 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:286:

5

```

Met Val Ala Met His Val Ser Ala Arg Arg Arg Ile Leu Ser Arg Cys
 1           5           10           15
Ala Ala Thr Ala Pro Ser Met Val Glu Pro Ser Ser Pro Gly Trp Trp
          20           25           30
10 Arg Ala Ser Leu Ser Arg Leu Thr Met Gln Ala Trp Tyr Val Arg Ala
    35           40           45
Arg Ala Arg Ala Phe Thr Arg Arg Arg Val Ser Ser Ser Asp Ser Arg
    50           55           60
Ala Ser Ser Ser Val Met Gly Ala Gly Lys Ser Ala Leu Thr Thr Ala
15 65           70           75           80
Arg Ala Ser Cys Ser Arg Gly Ser Xaa Ser Glu Gly Gly Ala Ala Ala
          85           90           95
Arg Ile Ile Ser Tyr Cys Cys Ser Ser Gly Arg Val Pro Gln Pro His
          100          105          110
20 Ser Thr Pro Ser Arg Asp Ala Ile Pro Glu His Arg Ser Ala Pro Ala
    115          120          125
Phe Pro His Pro Thr Pro Ser Gly Phe Ala Gly Ala Met Gly Thr Glu
    130          135          140
Asp Cys Asp His Glu Gly Arg Ser Val Ala Ala Pro Val Glu Val Met
25 145          150          155          160
Ala Leu Tyr Ala Thr Asp Gly Cys Val Ile Thr Ser Ser Leu Ala Leu
          165          170          175
Leu Thr Asn Cys Leu Leu Gly Ala Glu Pro Leu Tyr Ile Phe Ser Tyr
          180          185          190
30 Asp Ala Tyr Arg Pro Asp Ala Pro Asn Gly Pro Thr Gly Ala Pro Thr
    195          200          205
Glu Gln Glu Arg Phe Glu Gly Ser Arg Ala Leu Tyr Arg Asp Ala Gly
    210          215          220
Gln Gly Asp Ser Phe Arg Val Thr Phe Cys Leu Leu Gly Thr Glu Val
35 225          230          235          240
Gly Val Thr His His Pro Lys Gly Arg Trp Met Phe Val Cys Arg Phe
          245          250          255
Glu Arg Ala Asp Asp Val Ala Val Leu Gln Asp Ala Leu Gly Arg Gly
          260          265          270
40 Thr Pro Leu Leu Pro Ala His Ile Thr Ala Thr Leu Asp Leu Glu Ala
    275          280          285
Thr Phe Ala Leu His Ala Asn Ile Ile Met Ala Leu Thr Val Ala Ile
    290          295          300

```

Val His Asn Ala Pro Ala Arg Ile Gly Ser Gly Ser Thr Ala Pro Leu  
 305 310 315 320  
 Tyr Glu Pro Gly Glu Ser Met Arg Ser Val Val Gly Arg Met Ser Leu  
 325 330 335  
 5 Gly Gln Arg Gly Leu Thr Thr Leu Phe Val His His Glu Ala Arg Val  
 340 345 350  
 Leu Ala Ala Tyr Arg Arg Ala Tyr Tyr Gly Ser Ala Gln Ser Pro Phe  
 355 360 365  
 Trp Phe Leu Ser Lys Phe Gly Pro Asp Glu Lys Ser Leu Val Leu Ala  
 10 370 375 380  
 Ala Arg Tyr Tyr Val Leu Gln Ala Pro Arg Leu Gly Gly Ala Gly Ala  
 385 390 395 400  
 Thr Tyr Asp Leu Gln Ala Val Lys Asp Ile Cys Ala Thr Tyr Ala Ile  
 405 410 415  
 15 Pro His Asp Pro Arg Pro Asp Thr Leu Ser Ala Ala Ser Leu Thr Ser  
 420 425 430  
 Phe Ala Ala Ile Thr Arg Phe Cys Cys Thr Ser Gln Tyr Ser Arg Gly  
 435 440 445  
 Ala Ala Ala Ala Gly Phe Pro Leu Tyr Val Glu Arg Arg Ile Ala Ala  
 20 450 455 460  
 Asp Val Arg Glu Thr Gly Ala Leu Glu Lys Phe Ile Ala His Asp Arg  
 465 470 475 480  
 Ser Cys Leu Arg Val Ser Asp Arg Glu Phe Ile Thr Tyr Ile Tyr Leu  
 485 490 495  
 25 Ala His Phe Glu Cys Phe Ser Pro Pro Arg Leu Ala Thr His Leu Arg  
 500 505 510  
 Ala Val Thr Thr His Asp Pro Ser Pro Ala Ala Ser Thr Glu Gln Pro  
 515 520 525  
 Ser Pro Leu Gly Arg Glu Ala Val Glu Gln Phe Phe Arg His Val Arg  
 30 530 535 540  
 Ala Gln Leu Asn Ile Arg Glu Tyr Val Lys Gln Asn Val Thr Pro Arg  
 545 550 555 560  
 Glu Thr Ala Gly Asp Ala Ala Ala Ala Tyr Leu Arg Ala Arg Thr Tyr  
 565 570 575  
 35 Ala Pro Ala Ala Leu Thr Pro Ala Pro Ala Tyr Cys Gly Val Ala Asp  
 580 585 590  
 Ser Ser Thr Lys Met Met Gly Arg Leu Ala Glu Ala Glu Arg Leu Leu  
 595 600 605  
 Val Pro His Gly Trp Pro Ala Phe Ala Pro Thr Thr Pro Gly Asp Asp  
 40 610 615 620  
 Ala Gly Gly Gly Thr Ala Ala Pro Gln Thr Cys Gly Ile Val Lys Arg  
 625 630 635 640  
 Leu Leu Lys Leu Ala Ala Thr Glu Gln Gln Gly Thr Thr Pro Pro Ala

645 650 655  
 Ile Ala Ala Leu Met Gln Asp Ala Ser Val Gln Thr Pro Leu Pro Val  
 660 665 670  
 Tyr Arg Ile Thr Met Ser Pro Thr Gly Gln Ala Phe Ala Ala Ala Ala  
 5 675 680 685  
 Arg Asp Asp Trp Ala Arg Val Thr Arg Asp Ala Arg Pro Pro Glu Ala  
 690 695 700  
 Thr Val Val Ala Asp Ala Ala Ala Ala Pro Glu Pro Gly Ala Leu Gly  
 705 710 715 720  
 10 Arg Arg Leu Thr Arg Arg Ile Cys Arg Pro Ala Pro Pro Pro Gly Arg  
 725 730 735  
 Pro Gly Arg Arg Gly Pro Asp Val Arg Glu Pro Gln Arg Asp Leu Gln  
 740 745 750  
 Arg Arg Ala Gly Arg Tyr Glu His His Pro Gly Ser Gly His Arg Pro  
 15 755 760 765  
 Glu Gly Ala Arg Pro Leu Ser Pro Ala Pro Arg Gly Pro Gly Ser Leu  
 770 775 780

## (2) INFORMATION FOR SEQ ID NO:287:

20

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 789 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

25

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:287:

30

Met Tyr Ile Cys Arg Met Val Ala Met His Val Ser Ala Arg Arg Arg  
 1 5 10 15  
 Ile Leu Ser Arg Cys Ala Ala Thr Ala Pro Ser Met Val Glu Pro Ser  
 20 25 30  
 35 Ser Pro Gly Trp Trp Arg Ala Ser Leu Ser Arg Leu Thr Met Gln Ala  
 35 40 45  
 Trp Tyr Val Arg Ala Arg Ala Arg Ala Phe Thr Arg Arg Arg Val Ser  
 50 55 60  
 Ser Ser Asp Ser Arg Ala Ser Ser Ser Val Met Gly Ala Gly Lys Ser  
 40 65 70 75 80  
 Ala Leu Thr Thr Ala Arg Ala Ser Cys Ser Arg Gly Ser Xaa Ser Glu  
 85 90 95  
 Gly Gly Ala Ala Ala Arg Ile Ile Ser Tyr Cys Cys Ser Ser Gly Arg

	100		105		110
	Val Pro Gln Pro His Ser Thr Pro Ser Arg Asp Ala Ile Pro Glu His				
	115		120		125
	Arg Ser Ala Pro Ala Phe Pro His Pro Thr Pro Ser Gly Phe Ala Gly				
5	130		135		140
	Ala Met Gly Thr Glu Asp Cys Asp His Glu Gly Arg Ser Val Ala Ala				
	145		150		155
	Pro Val Glu Val Met Ala Leu Tyr Ala Thr Asp Gly Cys Val Ile Thr				
	165		170		175
10	Ser Ser Leu Ala Leu Leu Thr Asn Cys Leu Leu Gly Ala Glu Pro Leu				
	180		185		190
	Tyr Ile Phe Ser Tyr Asp Ala Tyr Arg Pro Asp Ala Pro Asn Gly Pro				
	195		200		205
	Thr Gly Ala Pro Thr Glu Gln Glu Arg Phe Glu Gly Ser Arg Ala Leu				
15	210		215		220
	Tyr Arg Asp Ala Gly Gln Gly Asp Ser Phe Arg Val Thr Phe Cys Leu				
	225		230		235
	Leu Gly Thr Glu Val Gly Val Thr His His Pro Lys Gly Arg Trp Met				
	245		250		255
20	Phe Val Cys Arg Phe Glu Arg Ala Asp Asp Val Ala Val Leu Gln Asp				
	260		265		270
	Ala Leu Gly Arg Gly Thr Pro Leu Leu Pro Ala His Ile Thr Ala Thr				
	275		280		285
	Leu Asp Leu Glu Ala Thr Phe Ala Leu His Ala Asn Ile Ile Met Ala				
25	290		295		300
	Leu Thr Val Ala Ile Val His Asn Ala Pro Ala Arg Ile Gly Ser Gly				
	305		310		315
	Ser Thr Ala Pro Leu Tyr Glu Pro Gly Glu Ser Met Arg Ser Val Val				
	325		330		335
30	Gly Arg Met Ser Leu Gly Gln Arg Gly Leu Thr Thr Leu Phe Val His				
	340		345		350
	His Glu Ala Arg Val Leu Ala Ala Tyr Arg Arg Ala Tyr Tyr Gly Ser				
	355		360		365
	Ala Gln Ser Pro Phe Trp Phe Leu Ser Lys Phe Gly Pro Asp Glu Lys				
35	370		375		380
	Ser Leu Val Leu Ala Ala Arg Tyr Tyr Val Leu Gln Ala Pro Arg Leu				
	385		390		395
	Gly Gly Ala Gly Ala Thr Tyr Asp Leu Gln Ala Val Lys Asp Ile Cys				
	405		410		415
40	Ala Thr Tyr Ala Ile Pro His Asp Pro Arg Pro Asp Thr Leu Ser Ala				
	420		425		430
	Ala Ser Leu Thr Ser Phe Ala Ala Ile Thr Arg Phe Cys Cys Thr Ser				
	435		440		445



Gln Tyr Ser Arg Gly Ala Ala Ala Gly Phe Pro Leu Tyr Val Glu  
 450 455 460  
 Arg Arg Ile Ala Ala Asp Val Arg Glu Thr Gly Ala Leu Glu Lys Phe  
 465 470 475 480  
 5 Ile Ala His Asp Arg Ser Cys Leu Arg Val Ser Asp Arg Glu Phe Ile  
 485 490 495  
 Thr Tyr Ile Tyr Leu Ala His Phe Glu Cys Phe Ser Pro Pro Arg Leu  
 500 505 510  
 Ala Thr His Leu Arg Ala Val Thr Thr His Asp Pro Ser Pro Ala Ala  
 10 515 520 525  
 Ser Thr Glu Gln Pro Ser Pro Leu Gly Arg Glu Ala Val Glu Gln Phe  
 530 535 540  
 Phe Arg His Val Arg Ala Gln Leu Asn Ile Arg Glu Tyr Val Lys Gln  
 545 550 555 560  
 15 Asn Val Thr Pro Arg Glu Thr Ala Gly Asp Ala Ala Ala Tyr Leu  
 565 570 575  
 Arg Ala Arg Thr Tyr Ala Pro Ala Ala Leu Thr Pro Ala Pro Ala Tyr  
 580 585 590  
 Cys Gly Val Ala Asp Ser Ser Thr Lys Met Met Gly Arg Leu Ala Glu  
 20 595 600 605  
 Ala Glu Arg Leu Leu Val Pro His Gly Trp Pro Ala Phe Ala Pro Thr  
 610 615 620  
 Thr Pro Gly Asp Asp Ala Gly Gly Gly Thr Ala Ala Pro Gln Thr Cys  
 625 630 635 640  
 25 Gly Ile Val Lys Arg Leu Leu Lys Leu Ala Ala Thr Glu Gln Gln Gly  
 645 650 655  
 Thr Thr Pro Pro Ala Ile Ala Ala Leu Met Gln Asp Ala Ser Val Gln  
 660 665 670  
 Thr Pro Leu Pro Val Tyr Arg Ile Thr Met Ser Pro Thr Gly Gln Ala  
 30 675 680 685  
 Phe Ala Ala Ala Ala Arg Asp Asp Trp Ala Arg Val Thr Arg Asp Ala  
 690 695 700  
 Arg Pro Pro Glu Ala Thr Val Val Ala Asp Ala Ala Ala Pro Glu  
 705 710 715 720  
 35 Pro Gly Ala Leu Gly Arg Arg Leu Thr Arg Arg Ile Cys Arg Pro Ala  
 725 730 735  
 Pro Pro Pro Gly Arg Pro Gly Arg Arg Gly Pro Asp Val Arg Glu Pro  
 740 745 750  
 Gln Arg Asp Leu Gln Arg Arg Ala Gly Arg Tyr Glu His His Pro Gly  
 40 755 760 765  
 Ser Gly His Arg Pro Glu Gly Ala Arg Pro Leu Ser Pro Ala Pro Arg  
 770 775 780  
 Gly Pro Gly Ser Leu

785

## (2) INFORMATION FOR SEQ ID NO:288:

## 5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 809 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:288:

15 Met Leu Arg Met Ala Trp Glu Thr Ser Thr Ser Ala Asp Leu Ser Ala  
 1 5 10 15  
 Ala Pro Thr Asp Met Tyr Ile Cys Arg Met Val Ala Met His Val Ser  
 20 25 30  
 Ala Arg Arg Arg Ile Leu Ser Arg Cys Ala Ala Thr Ala Pro Ser Met  
 20 35 40 45  
 Val Glu Pro Ser Ser Pro Gly Trp Trp Arg Ala Ser Leu Ser Arg Leu  
 50 55 60  
 Thr Met Gln Ala Trp Tyr Val Arg Ala Arg Ala Arg Ala Phe Thr Arg  
 65 70 75 80  
 25 Arg Arg Val Ser Ser Ser Asp Ser Arg Ala Ser Ser Ser Val Met Gly  
 85 90 95  
 Ala Gly Lys Ser Ala Leu Thr Thr Ala Arg Ala Ser Cys Ser Arg Gly  
 100 105 110  
 Ser Xaa Ser Glu Gly Gly Ala Ala Ala Arg Ile Ile Ser Tyr Cys Cys  
 115 120 125  
 30 Ser Ser Gly Arg Val Pro Gln Pro His Ser Thr Pro Ser Arg Asp Ala  
 130 135 140  
 Ile Pro Glu His Arg Ser Ala Pro Ala Phe Pro His Pro Thr Pro Ser  
 145 150 155 160  
 35 Gly Phe Ala Gly Ala Met Gly Thr Glu Asp Cys Asp His Glu Gly Arg  
 165 170 175  
 Ser Val Ala Ala Pro Val Glu Val Met Ala Leu Tyr Ala Thr Asp Gly  
 180 185 190  
 Cys Val Ile Thr Ser Ser Leu Ala Leu Leu Thr Asn Cys Leu Leu Gly  
 195 200 205  
 40 Ala Glu Pro Leu Tyr Ile Phe Ser Tyr Asp Ala Tyr Arg Pro Asp Ala  
 210 215 220  
 Pro Asn Gly Pro Thr Gly Ala Pro Thr Glu Gln Glu Arg Phe Glu Gly

707

	225		230		235		240									
	Ser	Arg	Ala	Leu	Tyr	Arg	Asp	Ala	Gly	Gln	Gly	Asp	Ser	Phe	Arg	Val
			245				250				255					
	Thr	Phe	Cys	Leu	Leu	Gly	Thr	Glu	Val	Gly	Val	Thr	His	His	Pro	Lys
5			260				265				270					
	Gly	Arg	Trp	Met	Phe	Val	Cys	Arg	Phe	Glu	Arg	Ala	Asp	Asp	Val	Ala
			275				280				285					
	Val	Leu	Gln	Asp	Ala	Leu	Gly	Arg	Gly	Thr	Pro	Leu	Leu	Pro	Ala	His
		290				295				300						
10	Ile	Thr	Ala	Thr	Leu	Asp	Leu	Glu	Ala	Thr	Phe	Ala	Leu	His	Ala	Asn
	305				310					315				320		
	Ile	Ile	Met	Ala	Leu	Thr	Val	Ala	Ile	Val	His	Asn	Ala	Pro	Ala	Arg
			325				330				335					
	Ile	Gly	Ser	Gly	Ser	Thr	Ala	Pro	Leu	Tyr	Glu	Pro	Gly	Glu	Ser	Met
15			340				345				350					
	Arg	Ser	Val	Val	Gly	Arg	Met	Ser	Leu	Gly	Gln	Arg	Gly	Leu	Thr	Thr
		355				360				365						
	Leu	Phe	Val	His	His	Glu	Ala	Arg	Val	Leu	Ala	Ala	Tyr	Arg	Arg	Ala
		370				375				380						
20	Tyr	Tyr	Gly	Ser	Ala	Gln	Ser	Pro	Phe	Trp	Phe	Leu	Ser	Lys	Phe	Gly
	385				390					395				400		
	Pro	Asp	Glu	Lys	Ser	Leu	Val	Leu	Ala	Ala	Arg	Tyr	Tyr	Val	Leu	Gln
			405				410			415						
	Ala	Pro	Arg	Leu	Gly	Gly	Ala	Gly	Ala	Thr	Tyr	Asp	Leu	Gln	Ala	Val
25			420				425			430						
	Lys	Asp	Ile	Cys	Ala	Thr	Tyr	Ala	Ile	Pro	His	Asp	Pro	Arg	Pro	Asp
		435				440				445						
	Thr	Leu	Ser	Ala	Ala	Ser	Leu	Thr	Ser	Phe	Ala	Ala	Ile	Thr	Arg	Phe
		450				455				460						
30	Cys	Cys	Thr	Ser	Gln	Tyr	Ser	Arg	Gly	Ala	Ala	Ala	Ala	Gly	Phe	Pro
	465				470					475				480		
	Leu	Tyr	Val	Glu	Arg	Arg	Ile	Ala	Ala	Asp	Val	Arg	Glu	Thr	Gly	Ala
			485				490			495						
	Leu	Glu	Lys	Phe	Ile	Ala	His	Asp	Arg	Ser	Cys	Leu	Arg	Val	Ser	Asp
35			500				505			510						
	Arg	Glu	Phe	Ile	Thr	Tyr	Ile	Tyr	Leu	Ala	His	Phe	Glu	Cys	Phe	Ser
		515				520				525						
	Pro	Pro	Arg	Leu	Ala	Thr	His	Leu	Arg	Ala	Val	Thr	Thr	His	Asp	Pro
		530				535				540						
40	Ser	Pro	Ala	Ala	Ser	Thr	Glu	Gln	Pro	Ser	Pro	Leu	Gly	Arg	Glu	Ala
	545				550					555				560		
	Val	Glu	Gln	Phe	Phe	Arg	His	Val	Arg	Ala	Gln	Leu	Asn	Ile	Arg	Glu
			565				570			575						

Tyr Val Lys Gln Asn Val Thr Pro Arg Glu Thr Ala Gly Asp Ala Ala  
 580 585 590  
 Ala Ala Tyr Leu Arg Ala Arg Thr Tyr Ala Pro Ala Ala Leu Thr Pro  
 595 600 605  
 5 Ala Pro Ala Tyr Cys Gly Val Ala Asp Ser Ser Thr Lys Met Met Gly  
 610 615 620  
 Arg Leu Ala Glu Ala Glu Arg Leu Leu Val Pro His Gly Trp Pro Ala  
 625 630 635 640  
 Phe Ala Pro Thr Thr Pro Gly Asp Asp Ala Gly Gly Gly Thr Ala Ala  
 10 645 650 655  
 Pro Gln Thr Cys Gly Ile Val Lys Arg Leu Leu Lys Leu Ala Ala Thr  
 660 665 670  
 Glu Gln Gln Gly Thr Thr Pro Pro Ala Ile Ala Ala Leu Met Gln Asp  
 675 680 685  
 15 Ala Ser Val Gln Thr Pro Leu Pro Val Tyr Arg Ile Thr Met Ser Pro  
 690 695 700  
 Thr Gly Gln Ala Phe Ala Ala Ala Ala Arg Asp Asp Trp Ala Arg Val  
 705 710 715 720  
 Thr Arg Asp Ala Arg Pro Pro Glu Ala Thr Val Val Ala Asp Ala Ala  
 20 725 730 735  
 Ala Ala Pro Glu Pro Gly Ala Leu Gly Arg Arg Leu Thr Arg Arg Ile  
 740 745 750  
 Cys Arg Pro Ala Pro Pro Pro Gly Arg Pro Gly Arg Arg Gly Pro Asp  
 755 760 765  
 25 Val Arg Glu Pro Gln Arg Asp Leu Gln Arg Arg Ala Gly Arg Tyr Glu  
 770 775 780  
 His His Pro Gly Ser Gly His Arg Pro Glu Gly Ala Arg Pro Leu Ser  
 785 790 795 800  
 Pro Ala Pro Arg Gly Pro Gly Ser Leu  
 30 805

## (2) INFORMATION FOR SEQ ID NO:289:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 816 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:289:

Met Thr Thr Ser Leu Ser Ala Met Leu Arg Met Ala Trp Glu Thr Ser  
 1 5 10 15  
 Thr Ser Ala Asp Leu Ser Ala Ala Pro Thr Asp Met Tyr Ile Cys Arg  
 20 25 30  
 5 Met Val Ala Met His Val Ser Ala Arg Arg Arg Ile Leu Ser Arg Cys  
 35 40 45  
 Ala Ala Thr Ala Pro Ser Met Val Glu Pro Ser Ser Pro Gly Trp Trp  
 50 55 60  
 Arg Ala Ser Leu Ser Arg Leu Thr Met Gln Ala Trp Tyr Val Arg Ala  
 10 65 70 75 80  
 Arg Ala Arg Ala Phe Thr Arg Arg Arg Val Ser Ser Ser Asp Ser Arg  
 85 90 95  
 Ala Ser Ser Ser Val Met Gly Ala Gly Lys Ser Ala Leu Thr Thr Ala  
 100 105 110  
 15 Arg Ala Ser Cys Ser Arg Gly Ser Xaa Ser Glu Gly Gly Ala Ala Ala  
 115 120 125  
 Arg Ile Ile Ser Tyr Cys Cys Ser Ser Gly Arg Val Pro Gln Pro His  
 130 135 140  
 Ser Thr Pro Ser Arg Asp Ala Ile Pro Glu His Arg Ser Ala Pro Ala  
 20 145 150 155 160  
 Phe Pro His Pro Thr Pro Ser Gly Phe Ala Gly Ala Met Gly Thr Glu  
 165 170 175  
 Asp Cys Asp His Glu Gly Arg Ser Val Ala Ala Pro Val Glu Val Met  
 180 185 190  
 25 Ala Leu Tyr Ala Thr Asp Gly Cys Val Ile Thr Ser Ser Leu Ala Leu  
 195 200 205  
 Leu Thr Asn Cys Leu Leu Gly Ala Glu Pro Leu Tyr Ile Phe Ser Tyr  
 210 215 220  
 Asp Ala Tyr Arg Pro Asp Ala Pro Asn Gly Pro Thr Gly Ala Pro Thr  
 30 225 230 235 240  
 Glu Gln Glu Arg Phe Glu Gly Ser Arg Ala Leu Tyr Arg Asp Ala Gly  
 245 250 255  
 Gln Gly Asp Ser Phe Arg Val Thr Phe Cys Leu Leu Gly Thr Glu Val  
 260 265 270  
 35 Gly Val Thr His His Pro Lys Gly Arg Trp Met Phe Val Cys Arg Phe  
 275 280 285  
 Glu Arg Ala Asp Asp Val Ala Val Leu Gln Asp Ala Leu Gly Arg Gly  
 290 295 300  
 Thr Pro Leu Leu Pro Ala His Ile Thr Ala Thr Leu Asp Leu Glu Ala  
 40 305 310 315 320  
 Thr Phe Ala Leu His Ala Asn Ile Ile Met Ala Leu Thr Val Ala Ile  
 325 330 335  
 Val His Asn Ala Pro Ala Arg Ile Gly Ser Gly Ser Thr Ala Pro Leu

	340	345	350
	Tyr Glu Pro Gly Glu Ser Met Arg Ser Val Val Gly Arg Met Ser Leu		
	355	360	365
	Gly Gln Arg Gly Leu Thr Thr Leu Phe Val His His Glu Ala Arg Val		
5	370	375	380
	Leu Ala Ala Tyr Arg Arg Ala Tyr Tyr Gly Ser Ala Gln Ser Pro Phe		
	385	390	395
	Trp Phe Leu Ser Lys Phe Gly Pro Asp Glu Lys Ser Leu Val Leu Ala		
	405	410	415
10	Ala Arg Tyr Tyr Val Leu Gln Ala Pro Arg Leu Gly Gly Ala Gly Ala		
	420	425	430
	Thr Tyr Asp Leu Gln Ala Val Lys Asp Ile Cys Ala Thr Tyr Ala Ile		
	435	440	445
	Pro His Asp Pro Arg Pro Asp Thr Leu Ser Ala Ala Ser Leu Thr Ser		
15	450	455	460
	Phe Ala Ala Ile Thr Arg Phe Cys Cys Thr Ser Gln Tyr Ser Arg Gly		
	465	470	475
	Ala Ala Ala Ala Gly Phe Pro Leu Tyr Val Glu Arg Arg Ile Ala Ala		
	485	490	495
20	Asp Val Arg Glu Thr Gly Ala Leu Glu Lys Phe Ile Ala His Asp Arg		
	500	505	510
	Ser Cys Leu Arg Val Ser Asp Arg Glu Phe Ile Thr Tyr Ile Tyr Leu		
	515	520	525
	Ala His Phe Glu Cys Phe Ser Pro Pro Arg Leu Ala Thr His Leu Arg		
25	530	535	540
	Ala Val Thr Thr His Asp Pro Ser Pro Ala Ala Ser Thr Glu Gln Pro		
	545	550	555
	Ser Pro Leu Gly Arg Glu Ala Val Glu Gln Phe Phe Arg His Val Arg		
	565	570	575
30	Ala Gln Leu Asn Ile Arg Glu Tyr Val Lys Gln Asn Val Thr Pro Arg		
	580	585	590
	Glu Thr Ala Gly Asp Ala Ala Ala Ala Tyr Leu Arg Ala Arg Thr Tyr		
	595	600	605
	Ala Pro Ala Ala Leu Thr Pro Ala Pro Ala Tyr Cys Gly Val Ala Ala		
35	610	615	620
	Asp Ser Ser Thr Lys Met Met Gly Arg Leu Ala Glu Ala Glu Arg Leu		
	625	630	635
	Leu Val Pro Gly Trp Pro Ala Phe Ala Pro Thr Thr Pro Gly Asp Asp		
	645	650	655
40	Ala Gly Gly Gly Thr Ala Ala Pro Gln Thr Cys Gly Ile Val Lys Arg		
	660	665	670
	Leu Leu Lys Leu Ala Ala Thr Glu Gln Gln Gly Thr Thr Pro Pro Ala		
	675	680	685

Ile Ala Ala Leu Met Gln Asp Ala Ser Val Gln Thr Pro Leu Pro Val  
 690 695 700  
 Tyr Arg Ile Thr Met Ser Pro Thr Gly Gln Ala Phe Ala Ala Ala Ala  
 705 710 715 720  
 5 Arg Asp Asp Trp Ala Arg Val Thr Arg Asp Ala Arg Pro Pro Glu Ala  
 725 730 735  
 Thr Val Val Ala Asp Ala Ala Ala Ala Pro Glu Pro Gly Ala Leu Gly  
 740 745 750  
 Arg Arg Leu Thr Arg Arg Ile Cys Arg Pro Ala Pro Pro Pro Gly Arg  
 10 755 760 765  
 Pro Gly Arg Arg Gly Pro Asp Val Arg Glu Pro Gln Arg Asp Leu Gln  
 770 775 780  
 Arg Arg Ala Gly Arg Tyr Glu His His Pro Gly Ser Gly His Arg Pro  
 785 790 795 800  
 15 Glu Gly Ala Arg Pro Leu Ser Pro Ala Pro Arg Gly Pro Gly Ser Leu  
 805 810 815

## (2) INFORMATION FOR SEQ ID NO:290:

## 20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 184 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

25

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:290:

30 Met Thr Thr Thr Pro Leu Ser Asn Leu Phe Leu Arg Ala Pro Asp Ile  
 1 5 10 15  
 Thr His Val Ala Pro Pro Tyr Cys Leu Asn Ala Thr Trp Gln Ala Glu  
 20 25 30  
 Asn Ala Leu His Thr Thr Lys Thr Asp Pro Ala Cys Leu Ala Ala Arg  
 35 35 40 45  
 Ser Tyr Leu Val Arg Ala Ser Cys Ser Thr Ser Gly Pro Ile His Cys  
 50 55 60  
 Phe Phe Phe Ala Val Tyr Lys Asp Ser Gln His Ser Leu Pro Leu Val  
 65 70 75 80  
 40 Thr Glu Leu Arg Asn Phe Ala Asp Leu Val Asn His Pro Pro Val Leu  
 85 90 95  
 Arg Glu Leu Glu Asp Lys Arg Gly Gly Arg Leu Arg Cys Thr Gly Pro  
 100 105 110  
 712

Phe Ser Cys Gly Thr Ile Lys Asp Val Ser Gly Asp Ala Gly Glu Tyr  
           115                          120                          125  
 Thr Ile Asn Gly Ile Val Tyr His Cys His Cys Arg Tyr Pro Phe Ser  
           130                          135                          140  
 5 Lys Thr Cys Trp Leu Gly Ala Ser Ala Ala Leu Gln His Leu Arg Ser  
    145                          150                          155                          160  
 Ile Ser Ser Ser Gly Thr Ala Ala Arg Ala Ala Glu Gln Arg Arg His  
                           165                          170                          175  
 Lys Ile Lys Ile Lys Ile Lys Val  
 10                          180

## (2) INFORMATION FOR SEQ ID NO:291:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 212 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:291:

25 Met Trp Gly Pro Gly Pro Ala Arg Phe Ile Ala Arg Pro Gly Thr His  
    1                          5                          10                          15  
 Gly Arg Arg Val Phe Thr Asp Pro Pro Pro Arg Asn Met Thr Thr Thr  
                           20                          25                          30  
 Pro Leu Ser Asn Leu Phe Leu Arg Ala Pro Asp Ile Thr His Val Ala  
                           35                          40                          45  
 30 Pro Pro Tyr Cys Leu Asn Ala Thr Trp Gln Ala Glu Asn Ala Leu His  
    50                          55                          60  
 Thr Thr Lys Thr Asp Pro Ala Cys Leu Ala Ala Arg Ser Tyr Leu Val  
    65                          70                          75                          80  
 Arg Ala Ser Cys Ser Thr Ser Gly Pro Ile His Cys Phe Phe Phe Ala  
                           85                          90                          95  
 35 Val Tyr Lys Asp Ser Gln His Ser Leu Pro Leu Val Thr Glu Leu Arg  
                           100                          105                          110  
 Asn Phe Ala Asp Leu Val Asn His Pro Pro Val Leu Arg Glu Leu Glu  
                           115                          120                          125  
 40 Asp Lys Arg Gly Gly Arg Leu Arg Cys Thr Gly Pro Phe Ser Cys Gly  
    130                          135                          140  
 Thr Ile Lys Asp Val Ser Gly Asp Ala Gly Glu Tyr Thr Ile Asn Gly  
    145                          150                          155                          160



[illegible]

10 (2) INFORMATION FOR SEQ ID NO:292:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 670 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:292:

	Met	Ala	Ala	Gln	Arg	Ala	Arg	Ala	Pro	Ala	Met	Arg	Thr	Arg	Gly	Gly
	1			5					10					15		
25	Asp	Ala	Ala	Leu	Cys	Ala	Pro	Glu	Asp	Gly	Trp	Val	Lys	Val	His	Pro
			20						25					30		
	Thr	Pro	Gly	Thr	Met	Leu	Phe	Arg	Glu	Ile	Leu	Leu	Gly	Gln	Met	Gly
			35						40					45		
	Tyr	Thr	Glu	Gly	Gln	Gly	Val	Tyr	Asn	Val	Val	Arg	Ser	Ser	Glu	Ala
		50					55						60			
30	Ala	Thr	Arg	Gln	Leu	Gln	Ala	Ala	Ile	Phe	His	Ala	Leu	Leu	Asn	Ala
	65					70					75				80	
	Thr	Tyr	Asp	Leu	Glu	Glu	Asp	Trp	Arg	Arg	His	Val	Val	Arg	Leu	Gln
				85					90					95		
	Pro	Gln	Arg	Leu	Val	Arg	Arg	Tyr	Arg	Asn	Ala	Arg	Glu	Gly	Asp	Ile
35				100					105					110		
	Ala	Gly	Val	Ala	Glu	Arg	Val	Phe	Asp	Thr	Trp	Arg	Cys	Thr	Leu	Arg
			115					120					125			
	Thr	Thr	Leu	Leu	Asp	Phe	Ala	His	Gly	Val	Val	Asp	Cys	Phe	Ala	Pro
		130					135					140				
40	Gly	Gly	Pro	Ser	Gly	Pro	Thr	Ser	Phe	Pro	Lys	Tyr	Ile	Asp	Trp	Leu
	145					150					155					160
	Thr	Cys	Leu	Gly	Leu	Val	Pro	Ile	Leu	Arg	Lys	Thr	Arg	Glu	Gly	Glu
				165						170					175	

Ala Thr Gln Arg Leu Gly Ala Phe Leu Arg Gln His Thr Leu Pro Arg  
 180 185 190  
 Gln Leu Ala Thr Val Ala Gly Ala Ala Glu Arg Ala Gly Pro Gly Leu  
 195 200 205  
 5 Leu Glu Leu Ala Val Ala Phe Asp Ser Thr Arg Met Ala Glu Tyr Asp  
 210 215 220  
 Arg Val His Ile Tyr Tyr Asn His Arg Arg Gly Glu Trp Leu Val Arg  
 225 230 235 240  
 Asp Pro Val Ser Gly Gln Arg Gly Glu Cys Leu Val Leu Cys Pro Pro  
 10 245 250 255  
 Leu Trp Thr Gly Asp Arg Leu Val Phe Asp Ser Pro Val Gln Arg Leu  
 260 265 270  
 Cys Pro Glu Ile Val Ala Cys His Ala Leu Arg Glu His Ala His Ile  
 275 280 285  
 15 Cys Arg Leu Arg Asn Thr Ala Ser Val Lys Val Leu Leu Gly Arg Lys  
 290 295 300  
 Ser Asp Ser Gly Val Ala Gly Ala Ala Ala Arg Val Val Asn Lys Ala  
 305 310 315 320  
 Leu Gly Glu Asp Asp Glu Thr Lys Ala Gly Ser Ala Ala Ser Arg Leu  
 20 325 330 335  
 Val Arg Leu Ile Ile Met Lys Gly Met Arg His Val Gly Asp Ile Asn  
 340 345 350  
 Asp Thr Val Arg Ala Tyr Leu Asp Glu Ala Gly Gly His Leu Ile Asp  
 355 360 365  
 25 Thr Pro Ala Val Asp His Thr Leu Pro Gly Phe Gly Lys Gly Gly Thr  
 370 375 380  
 Gly Arg Gly Ser Ala Ala Gln Asp Pro Gly Ala Arg Pro Gln Gln Leu  
 385 390 395 400  
 Arg Gln Ala Phe Gln Thr Ala Val Val Asn Asn Ile Asn Gly Met Leu  
 30 405 410 415  
 Glu Gly Tyr Ile Asn Asn Leu Phe Gly Thr Ile Glu Arg Leu Arg Glu  
 420 425 430  
 Thr Asn Ala Gly Leu Ala Thr Gln Leu Gln Ala Arg Asp Arg Glu Leu  
 435 440 445  
 35 Arg Arg Ala Gln Ala Gly Ala Leu Glu Arg Glu Gln Arg Ala Ala Asp  
 450 455 460  
 Arg Ala Ala Gly Gly Gly Ala Gly Arg Pro Ala Glu Ala Asp Leu Leu  
 465 470 475 480  
 Arg Ala Asp Tyr Asp Ile Ile Asp Val Ser Lys Ser Met Asp Asp Asp  
 40 485 490 495  
 Thr Tyr Val Ala Asn Ser Phe Gln His Gln Tyr Ile Pro Ala Tyr Gly  
 500 505 510  
 Gln Asp Leu Glu Arg Leu Ser Arg Leu Trp Glu His Glu Leu Val Arg  
 715

515                      520                      525  
 Cys Phe Lys Ile Leu Arg His Arg Asn Asn Gln Gly Gln Glu Thr Ser  
      530                      535                      540  
 Ile Ser Tyr Ser Ser Gly Ala Ile Ala Ser Phe Val Ala Pro Tyr Phe  
 5    545                      550                      555                      560  
 Glu Tyr Val Leu Arg Ala Pro Arg Ala Gly Ala Leu Ile Thr Gly Ser  
                          565                      570                      575  
 Asp Val Ile Leu Gly Glu Glu Glu Leu Trp Glu Ala Val Phe Lys Lys  
                          580                      585                      590  
 10    Thr Arg Leu Gln Thr Tyr Leu Thr Asp Val Ala Ala Leu Phe Val Ala  
                          595                      600                      605  
 Asp Val Gln His Ala Ala Leu Pro Arg Pro Pro Ser Pro Thr Pro Ala  
                          610                      615                      620  
 Asp Phe Arg Ala Ser Asp Arg Gly Gly Ser Arg Ser Arg Thr Arg Thr  
 15    625                      630                      635                      640  
 Arg Ser Arg Ser Pro Gly Arg Thr Pro Arg Gly Ala Pro Asp Gln Gly  
                          645                      650                      655  
 Trp Gly Val Glu Arg Arg Asp Gly Arg Pro His Ala Arg Arg  
                          660                      665                      670  
 20

## (2) INFORMATION FOR SEQ ID NO:293:

## (i) SEQUENCE CHARACTERISTICS:

- 25    (A) LENGTH: 710 amino acids  
       (B) TYPE: amino acid  
       (C) STRANDEDNESS: single  
       (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:293:

Met Asp Val Lys Phe Lys Asn Ala Ser Ser Leu Asn Arg Thr Ala Gly  
   1                      5                      10                      15  
 35    Leu Ala Pro Gly Cys Cys Gly Gly Gly Pro Gly Ala Arg Thr Ser Arg  
                          20                      25                      30  
 Glu Pro Ser Pro Pro Asp Ala Ala Met Ala Ala Gln Arg Ala Arg Ala  
                          35                      40                      45  
 Pro Ala Met Arg Thr Arg Gly Gly Asp Ala Ala Leu Cys Ala Pro Glu  
 40    50                      55                      60  
 Asp Gly Trp Val Lys Val His Pro Thr Pro Gly Thr Met Leu Phe Arg  
   65                      70                      75                      80  
 Glu Ile Leu Leu Gly Gln Met Gly Tyr Thr Glu Gly Gln Gly Val Tyr

		85		90		95	
		Asn Val Val Arg Ser Ser Glu Ala Ala Thr Arg Gln Leu Gln Ala Ala					
		100		105		110	
		Ile Phe His Ala Leu Leu Asn Ala Thr Tyr Asp Leu Glu Glu Asp Trp					
5		115		120		125	
		Arg Arg His Val Val Arg Leu Gln Pro Gln Arg Leu Val Arg Arg Tyr					
		130		135		140	
		Arg Asn Ala Arg Glu Gly Asp Ile Ala Gly Val Ala Glu Arg Val Phe					
		145		150		155	160
10		Asp Thr Trp Arg Cys Thr Leu Arg Thr Thr Leu Leu Asp Phe Ala His					
		165		170		175	
		Gly Val Val Asp Cys Phe Ala Pro Gly Gly Pro Ser Gly Pro Thr Ser					
		180		185		190	
		Phe Pro Lys Tyr Ile Asp Trp Leu Thr Cys Leu Gly Leu Val Pro Ile					
15		195		200		205	
		Leu Arg Lys Thr Arg Glu Gly Glu Ala Thr Gln Arg Leu Gly Ala Phe					
		210		215		220	
		Leu Arg Gln His Thr Leu Pro Arg Gln Leu Ala Thr Val Ala Gly Ala					
		225		230		235	240
20		Ala Glu Arg Ala Gly Pro Gly Leu Leu Glu Leu Ala Val Ala Phe Asp					
		245		250		255	
		Ser Thr Arg Met Ala Glu Tyr Asp Arg Val His Ile Tyr Tyr Asn His					
		260		265		270	
		Arg Arg Gly Glu Trp Leu Val Arg Asp Pro Val Ser Gly Gln Arg Gly					
25		275		280		285	
		Glu Cys Leu Val Leu Cys Pro Pro Leu Trp Thr Gly Asp Arg Leu Val					
		290		295		300	
		Phe Asp Ser Pro Val Gln Arg Leu Cys Pro Glu Ile Val Ala Cys His					
		305		310		315	320
30		Ala Leu Arg Glu His Ala His Ile Cys Arg Leu Arg Asn Thr Ala Ser					
		325		330		335	
		Val Lys Val Leu Leu Gly Arg Lys Ser Asp Ser Gly Val Ala Gly Ala					
		340		345		350	
		Ala Arg Val Val Asn Lys Ala Leu Gly Glu Asp Asp Glu Thr Lys Ala					
35		355		360		365	
		Gly Ser Ala Ala Ser Arg Leu Val Arg Leu Ile Ile Asn Met Lys Gly					
		370		375		380	
		Met Arg His Val Gly Asp Ile Asn Asp Thr Val Arg Ala Tyr Leu Asp					
		385		390		395	400
40		Glu Ala Gly Gly His Leu Ile Asp Thr Pro Ala Val Asp His Thr Leu					
		405		410		415	
		Pro Gly Phe Gly Lys Gly Gly Thr Gly Arg Gly Ser Ala Ala Gln Asp					
		420		425		430	

Pro Gly Ala Arg Pro Gln Gln Leu Arg Gln Ala Phe Gln Thr Ala Val  
                   435                                  440                                  445  
 Val Asn Asn Ile Asn Gly Met Leu Glu Gly Tyr Ile Asn Asn Leu Phe  
                   450                                  455                                  460  
 5 Gly Thr Ile Glu Arg Leu Arg Glu Thr Asn Ala Gly Leu Ala Thr Gln  
                   465                                  470                                  475                                  480  
 Leu Gln Ala Arg Asp Arg Glu Leu Arg Arg Ala Gln Ala Gly Ala Leu  
                                   485                                  490                                  495  
 Glu Arg Glu Gln Arg Ala Ala Asp Arg Ala Ala Gly Gly Gly Ala Gly  
 10                                   500                                  505                                  510  
 Arg Pro Ala Glu Ala Asp Leu Leu Arg Ala Asp Tyr Asp Ile Ile Asp  
                   515                                  520                                  525  
 Val Ser Lys Ser Met Asp Asp Asp Thr Tyr Val Ala Asn Ser Phe Gln  
                   530                                  535                                  540  
 15 His Gln Tyr Ile Pro Ala Tyr Gly Gln Asp Leu Glu Arg Leu Ser Arg  
                   545                                  550                                  555                                  560  
 Leu Trp Glu His Glu Leu Val Arg Cys Phe Lys Ile Leu Arg His Arg  
                                   565                                  570                                  575  
 Asn Asn Gln Gly Gln Glu Thr Ser Ile Ser Tyr Ser Ser Gly Ala Ile  
 20                                   580                                  585                                  590  
 Ala Ser Phe Val Ala Pro Tyr Phe Glu Tyr Val Leu Arg Ala Pro Arg  
                   595                                  600                                  605  
 Ala Gly Ala Leu Ile Thr Gly Ser Asp Val Ile Leu Gly Glu Glu Glu  
                   610                                  615                                  620  
 25 Leu Trp Glu Ala Val Phe Lys Lys Thr Arg Leu Gln Thr Tyr Leu Thr  
                   625                                  630                                  635                                  640  
 Asp Val Ala Ala Leu Phe Val Ala Asp Val Gln His Ala Ala Leu Pro  
                                   645                                  650                                  655  
 Arg Pro Pro Ser Pro Thr Pro Ala Asp Phe Arg Ala Ser Asp Arg Gly  
 30                                   660                                  665                                  670  
 Gly Ser Arg Ser Arg Thr Arg Thr Arg Ser Arg Ser Pro Gly Arg Thr  
                   675                                  680                                  685  
 Pro Arg Gly Ala Pro Asp Gln Gly Trp Gly Val Glu Arg Arg Asp Gly  
                   690                                  695                                  700  
 35 Arg Pro His Ala Arg Arg  
                   705                                  710

## (2) INFORMATION FOR SEQ ID NO:294:

- 40 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 720 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:294:

	Met	Arg	Ala	Met	Ile	Gly	Trp	Thr	Pro	Cys	Met	Asp	Val	Lys	Phe	Lys
	1				5				10					15		
	Asn	Ala	Ser	Ser	Leu	Asn	Arg	Thr	Ala	Gly	Leu	Ala	Pro	Gly	Cys	Cys
10				20					25					30		
	Gly	Gly	Gly	Pro	Gly	Ala	Arg	Thr	Ser	Arg	Glu	Pro	Ser	Pro	Pro	Asp
				35				40						45		
	Ala	Ala	Met	Ala	Ala	Gln	Arg	Ala	Arg	Ala	Pro	Ala	Met	Arg	Thr	Arg
				50				55						60		
15	Gly	Gly	Asp	Ala	Ala	Leu	Cys	Ala	Pro	Glu	Asp	Gly	Trp	Val	Lys	Val
	65					70					75				80	
	His	Pro	Thr	Pro	Gly	Thr	Met	Leu	Phe	Arg	Glu	Ile	Leu	Leu	Gly	Gln
					85					90					95	
	Met	Gly	Tyr	Thr	Glu	Gly	Gln	Gly	Val	Tyr	Asn	Val	Val	Arg	Ser	Ser
20				100					105					110		
	Glu	Ala	Ala	Thr	Arg	Gln	Leu	Gln	Ala	Ala	Ile	Phe	His	Ala	Leu	Leu
				115				120						125		
	Asn	Ala	Thr	Tyr	Asp	Leu	Glu	Glu	Asp	Trp	Arg	Arg	His	Val	Val	Arg
				130				135						140		
25	Leu	Gln	Pro	Gln	Arg	Leu	Val	Arg	Arg	Tyr	Arg	Asn	Ala	Arg	Glu	Gly
	145					150					155				160	
	Asp	Ile	Ala	Gly	Val	Ala	Glu	Arg	Val	Phe	Asp	Thr	Trp	Arg	Cys	Thr
				165						170				175		
	Leu	Arg	Thr	Thr	Leu	Leu	Asp	Phe	Ala	His	Gly	Val	Val	Asp	Cys	Phe
30				180					185					190		
	Ala	Pro	Gly	Gly	Pro	Ser	Gly	Pro	Thr	Ser	Phe	Pro	Lys	Tyr	Ile	Asp
				195				200						205		
	Trp	Leu	Thr	Cys	Leu	Gly	Leu	Val	Pro	Ile	Leu	Arg	Lys	Thr	Arg	Glu
				210				215						220		
35	Gly	Glu	Ala	Thr	Gln	Arg	Leu	Gly	Ala	Phe	Leu	Arg	Gln	His	Thr	Leu
	225					230					235				240	
	Pro	Arg	Gln	Leu	Ala	Thr	Val	Ala	Gly	Ala	Ala	Glu	Arg	Ala	Gly	Pro
					245					250					255	
	Gly	Leu	Leu	Glu	Leu	Ala	Val	Ala	Phe	Asp	Ser	Thr	Arg	Met	Ala	Glu
40				260					265					270		
	Tyr	Asp	Arg	Val	His	Ile	Tyr	Tyr	Asn	His	Arg	Arg	Gly	Glu	Trp	Leu
				275					280					285		
	Val	Arg	Asp	Pro	Val	Ser	Gly	Gln	Arg	Gly	Glu	Cys	Leu	Val	Leu	Cys

	290		295		300											
	Pro	Pro	Leu	Trp	Thr	Gly	Asp	Arg	Leu	Val	Phe	Asp	Ser	Pro	Val	Gln
	305					310					315				320	
	Arg	Leu	Cys	Pro	Glu	Ile	Val	Ala	Cys	His	Ala	Leu	Arg	Glu	His	Ala
5					325					330				335		
	His	Ile	Cys	Arg	Leu	Arg	Asn	Thr	Ala	Ser	Val	Lys	Val	Leu	Leu	Gly
				340					345					350		
	Arg	Lys	Ser	Asp	Ser	Gly	Val	Ala	Gly	Ala	Ala	Arg	Val	Val	Asn	Lys
		355					360						365			
10	Ala	Leu	Gly	Glu	Asp	Asp	Glu	Thr	Lys	Ala	Gly	Ser	Ala	Ala	Ser	Arg
	370						375					380				
	Leu	Val	Arg	Leu	Ile	Ile	Asn	Met	Lys	Gly	Met	Arg	His	Val	Gly	Asp
	385				390					395					400	
	Ile	Asn	Asp	Thr	Val	Arg	Ala	Tyr	Leu	Asp	Glu	Ala	Gly	Gly	His	Leu
15				405					410					415		
	Ile	Asp	Thr	Pro	Ala	Val	Asp	His	Thr	Leu	Pro	Gly	Phe	Gly	Lys	Gly
			420						425					430		
	Gly	Thr	Gly	Arg	Gly	Ser	Ala	Ala	Gln	Asp	Pro	Gly	Ala	Arg	Pro	Gln
		435					440						445			
20	Gln	Leu	Arg	Gln	Ala	Phe	Gln	Thr	Ala	Val	Val	Asn	Asn	Ile	Asn	Gly
	450						455					460				
	Met	Leu	Glu	Gly	Tyr	Ile	Asn	Asn	Leu	Phe	Gly	Thr	Ile	Glu	Arg	Leu
	465				470					475				480		
	Arg	Glu	Thr	Asn	Ala	Gly	Leu	Ala	Thr	Gln	Leu	Gln	Ala	Arg	Asp	Arg
25				485					490					495		
	Glu	Leu	Arg	Arg	Ala	Gln	Ala	Gly	Ala	Leu	Glu	Arg	Glu	Gln	Arg	Ala
		500						505					510			
	Ala	Asp	Arg	Ala	Ala	Gly	Gly	Gly	Ala	Gly	Arg	Pro	Ala	Glu	Ala	Asp
		515					520					525				
30	Leu	Leu	Arg	Ala	Asp	Tyr	Asp	Ile	Ile	Asp	Val	Ser	Lys	Ser	Met	Asp
	530						535					540				
	Asp	Asp	Thr	Tyr	Val	Ala	Asn	Ser	Phe	Gln	His	Gln	Tyr	Ile	Pro	Ala
	545				550					555				560		
	Tyr	Gly	Gln	Asp	Leu	Glu	Arg	Leu	Ser	Arg	Leu	Trp	Glu	His	Glu	Leu
35				565					570					575		
	Val	Arg	Cys	Phe	Lys	Ile	Leu	Arg	His	Arg	Asn	Asn	Gln	Gly	Gln	Glu
		580						585					590			
	Thr	Ser	Ile	Ser	Tyr	Ser	Ser	Gly	Ala	Ile	Ala	Ser	Phe	Val	Ala	Pro
		595					600					605				
40	Tyr	Phe	Glu	Tyr	Val	Leu	Arg	Ala	Pro	Arg	Ala	Gly	Ala	Leu	Ile	Thr
	610						615					620				
	Gly	Ser	Asp	Val	Ile	Leu	Gly	Glu	Glu	Glu	Leu	Trp	Glu	Ala	Val	Phe
	625				630				635					640		

Lys Lys Thr Arg Leu Gln Thr Tyr Leu Thr Asp Val Ala Ala Leu Phe  
                                 645                                650                                655  
 Val Ala Asp Val Gln His Ala Ala Leu Pro Arg Pro Pro Ser Pro Thr  
                                 660                                665                                670  
 5 Pro Ala Asp Phe Arg Ala Ser Asp Arg Gly Gly Ser Arg Ser Arg Thr  
                                 675                                680                                685  
 Arg Thr Arg Ser Arg Ser Pro Gly Arg Thr Pro Arg Gly Ala Pro Asp  
                                 690                                695                                700  
 Gln Gly Trp Gly Val Glu Arg Arg Asp Gly Arg Pro His Ala Arg Arg  
 10 705                                710                                715                                720

## (2) INFORMATION FOR SEQ ID NO:295:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 763 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 20 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:295:

Met Arg Tyr Ala Ala Asn Gly Asn Ser Arg Ser Gly Arg Pro Val Gly  
 25 1                                5                                10                                15  
 Thr Ser Lys Ala Ala Thr Ser Arg Asn His Cys Arg Arg Gly Thr Cys  
                                 20                                25                                30  
 Val Thr Ser Ser Cys Cys Cys Glu Ser Ser Arg Met Arg Ala Met Ile  
                                 35                                40                                45  
 30 Gly Trp Thr Pro Cys Met Asp Val Lys Phe Lys Asn Ala Ser Ser Leu  
                                 50                                55                                60  
 Asn Arg Thr Ala Gly Leu Ala Pro Gly Cys Cys Gly Gly Gly Pro Gly  
                                 65                                70                                75                                80  
 Ala Arg Thr Ser Arg Glu Pro Ser Pro Pro Asp Ala Ala Met Ala Ala  
 35                                85                                90                                95  
 Gln Arg Ala Arg Ala Pro Ala Met Arg Thr Arg Gly Gly Asp Ala Ala  
                                 100                                105                                110  
 Leu Cys Ala Pro Glu Asp Gly Trp Val Lys Val His Pro Thr Pro Gly  
                                 115                                120                                125  
 40 Thr Met Leu Phe Arg Glu Ile Leu Leu Gly Gln Met Gly Tyr Thr Glu  
                                 130                                135                                140  
 Gly Gln Gly Val Tyr Asn Val Val Arg Ser Ser Glu Ala Ala Thr Arg  
 145                                150                                155                                160



Gln Leu Gln Ala Ala Ile Phe His Ala Leu Leu Asn Ala Thr Tyr Asp  
 165 170 175  
 Leu Glu Glu Asp Trp Arg Arg His Val Val Arg Leu Gln Pro Gln Arg  
 180 185 190  
 5 Leu Val Arg Arg Tyr Arg Asn Ala Arg Glu Gly Asp Ile Ala Gly Val  
 195 200 205  
 Ala Glu Arg Val Phe Asp Thr Trp Arg Cys Thr Leu Arg Thr Thr Leu  
 210 215 220  
 Leu Asp Phe Ala His Gly Val Val Asp Cys Phe Ala Pro Gly Gly Pro  
 10 225 230 235 240  
 Ser Gly Pro Thr Ser Phe Pro Lys Tyr Ile Asp Trp Leu Thr Cys Leu  
 245 250 255  
 Gly Leu Val Pro Ile Leu Arg Lys Thr Arg Glu Gly Glu Ala Thr Gln  
 260 265 270  
 15 Arg Leu Gly Ala Phe Leu Arg Gln His Thr Leu Pro Arg Gln Leu Ala  
 275 280 285  
 Thr Val Ala Gly Ala Ala Glu Arg Ala Gly Pro Gly Leu Leu Glu Leu  
 290 295 300  
 Ala Val Ala Phe Asp Ser Thr Arg Met Ala Glu Tyr Asp Arg Val His  
 20 305 310 315 320  
 Ile Tyr Tyr Asn His Arg Arg Gly Glu Trp Leu Val Arg Asp Pro Val  
 325 330 335  
 Ser Gly Gln Arg Gly Glu Cys Leu Val Leu Cys Pro Pro Leu Trp Thr  
 340 345 350  
 25 Gly Asp Arg Leu Val Phe Asp Ser Pro Val Gln Arg Leu Cys Pro Glu  
 355 360 365  
 Ile Val Ala Cys His Ala Leu Arg Glu His Ala His Ile Cys Arg Leu  
 370 375 380  
 Arg Asn Thr Ala Ser Val Lys Val Leu Leu Gly Arg Lys Ser Asp Ser  
 30 385 390 395 400  
 Gly Val Ala Gly Ala Ala Arg Val Val Asn Lys Ala Leu Gly Glu Asp  
 405 410 415  
 Asp Glu Thr Lys Ala Gly Ser Ala Ala Ser Arg Leu Val Arg Leu Ile  
 420 425 430  
 35 Ile Asn Met Lys Gly Met Arg His Val Gly Asp Ile Asn Asp Thr Val  
 435 440 445  
 Arg Ala Tyr Leu Asp Glu Ala Gly Gly His Leu Ile Asp Thr Pro Ala  
 450 455 460  
 Val Asp His Thr Leu Pro Gly Phe Gly Lys Gly Gly Thr Gly Arg Gly  
 40 465 470 475 480  
 Ser Ala Ala Gln Asp Pro Gly Ala Arg Pro Gln Gln Leu Arg Gln Ala  
 485 490 495  
 Phe Gln Thr Ala Val Val Asn Asn Ile Asn Gly Met Leu Glu Gly Tyr

	500		505		510
	Ile Asn Asn Leu Phe Gly Thr	Ile Glu Arg Leu Arg Glu Thr Asn Ala			
	515	520	525		
	Gly Leu Ala Thr Gln Leu Gln Ala Arg Asp Arg Glu Leu Arg Arg Ala				
5	530	535	540		
	Gln Ala Gly Ala Leu Glu Arg Glu Gln Arg Ala Ala Asp Arg Ala Ala				
	545	550	555	560	
	Gly Gly Gly Ala Gly Arg Pro Ala Glu Ala Asp Leu Leu Arg Ala Asp				
	565	570	575		
10	Tyr Asp Ile Ile Asp Val Ser Lys Ser Met Asp Asp Asp Thr Tyr Val				
	580	585	590		
	Ala Asn Ser Phe Gln His Gln Tyr Ile Pro Ala Tyr Gly Gln Asp Leu				
	595	600	605		
	Glu Arg Leu Ser Arg Leu Trp Glu His Glu Leu Val Arg Cys Phe Lys				
15	610	615	620		
	Ile Leu Arg His Arg Asn Asn Gln Gly Gln Glu Thr Ser Ile Ser Tyr				
	625	630	635	640	
	Ser Ser Gly Ala Ile Ala Ser Phe Val Ala Pro Tyr Phe Glu Tyr Val				
	645	650	655		
20	Leu Arg Ala Pro Arg Ala Gly Ala Leu Ile Thr Gly Ser Asp Val Ile				
	660	665	670		
	Leu Gly Glu Glu Glu Leu Trp Glu Ala Val Phe Lys Lys Thr Arg Leu				
	675	680	685		
	Gln Thr Tyr Leu Thr Asp Val Ala Ala Leu Phe Val Ala Asp Val Gln				
25	690	695	700		
	His Ala Ala Leu Pro Arg Pro Pro Ser Pro Thr Pro Ala Asp Phe Arg				
	705	710	715	720	
	Ala Ser Asp Arg Gly Gly Ser Arg Ser Arg Thr Arg Thr Arg Ser Arg				
	725	730	735		
30	Ser Pro Gly Arg Thr Pro Arg Gly Ala Pro Asp Gln Gly Trp Gly Val				
	740	745	750		
	Glu Arg Arg Asp Gly Arg Pro His Ala Arg Arg				
	755	760			

35 (2) INFORMATION FOR SEQ ID NO:296:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 414 amino acids

(B) TYPE: amino acid

40 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:296:

	Met	Ala	Asp	Ile	Pro	Pro	Asp	Pro	Pro	Ala	Leu	Asn	Thr	Thr	Pro	Ala
5	1				5					10					15	
	Asn	His	Ala	Pro	Pro	Ser	Pro	Pro	Pro	Gly	Ser	Arg	Lys	Arg	Arg	Arg
				20					25				30			
	Pro	Val	Leu	Pro	Ser	Ser	Ser	Glu	Ser	Glu	Gly	Lys	Pro	Asp	Thr	Glu
			35					40					45			
10	Ser	Glu	Ser	Ser	Ser	Thr	Glu	Ser	Ser	Glu	Asp	Glu	Ala	Gly	Asp	Leu
		50					55					60				
	Arg	Gly	Gly	Arg	Arg	Arg	Ser	Pro	Arg	Glu	Leu	Gly	Gly	Arg	Tyr	Phe
	65					70				75					80	
	Leu	Asp	Leu	Ser	Ala	Glu	Ser	Thr	Thr	Gly	Thr	Glu	Ser	Glu	Gly	Thr
15					85					90					95	
	Gly	Pro	Ser	Asp	Asp	Asp	Asp	Asp	Ala	Ser	Asp	Gly	Trp	Leu	Val	
				100				105				110				
	Asp	Thr	Pro	Pro	Arg	Lys	Ser	Lys	Arg	Pro	Arg	Ile	Asn	Leu	Arg	Leu
				115				120				125				
20	Thr	Ser	Ser	Pro	Asp	Arg	Arg	Ala	Gly	Val	Val	Phe	Pro	Glu	Val	Trp
				130			135					140				
	Arg	Asn	Asp	Arg	Pro	Ile	Arg	Ala	Ala	Gln	Pro	Gln	Ala	Pro	Ala	Gln
	145					150				155					160	
	Ser	Ser	Gly	Asp	Arg	Ala	Ala	Ala	Pro	Arg	Arg	Ser	Ala	Arg	Gln	Ala
25					165					170					175	
	Gln	Met	Arg	Ser	Gly	Ala	Ala	Trp	Thr	Leu	Asp	Leu	His	Tyr	Ile	Arg
				180				185				190				
	Gln	Cys	Val	Asn	Gln	Leu	Phe	Arg	Ile	Leu	Arg	Ala	Ala	Pro	Asn	Pro
			195					200				205				
30	Pro	Gly	Ser	Ala	Asn	Arg	Leu	Arg	His	Leu	Val	Arg	Asp	Cys	Tyr	Leu
		210				215						220				
	Met	Gly	Tyr	Cys	Arg	Thr	Arg	Leu	Gly	Pro	Arg	Thr	Trp	Gly	Arg	Leu
	225					230				235				240		
	Leu	Gln	Ile	Ser	Gly	Gly	Thr	Trp	Asp	Val	Arg	Leu	Arg	Asn	Ala	Ile
35					245					250				255		
	Arg	Glu	Val	Glu	Ala	Arg	Phe	Glu	Pro	Ala	Ala	Glu	Pro	Val	Cys	Glu
				260				265				270				
	Leu	Pro	Cys	Leu	Asn	Ala	Arg	Arg	Tyr	Gly	Pro	Glu	Cys	Asp	Val	Gly
			275				280					285				
40	Asn	Leu	Glu	Thr	Asn	Gly	Gly	Ser	Thr	Ser	Asp	Asp	Glu	Ile	Ser	Asp
		290				295						300				
	Ala	Thr	Asp	Ser	Asp	Asp	Thr	Leu	Ala	Ser	His	Ser	Asp	Thr	Glu	Gly
	305					310				315					320	

(2) INFORMATION FOR SEQ ID NO:297:

(i) SEQUENCE CHARACTERISTICS:

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:297:

Met	Gly	Arg	Leu	Arg	Asn	Ala	Pro	Glu	Ser	Leu	Thr	Tyr	Met	Phe	Cys
1				5					10					15	
Ala	Ala	Ile	Arg	Val	Ala	Pro	Val	Thr	Thr	Gln	Ser	Arg	Thr	Ser	Leu
		20						25					30		
Arg	Val	Cys	Thr	His	Val	Leu	Phe	Pro	Asp	Pro	Ala	Leu	Pro	Val	Met
		35					40						45		
Arg	Tyr	Ala	Ala	Asn	Gly	Asn	Ser	Arg	Ser	Gly	Arg	Pro	Val	Gly	Thr
	50					55					60				
Ser	Lys	Ala	Ala	Thr	Ser	Arg	Asn	His	Cys	Arg	Arg	Gly	Thr	Cys	Val
65				70					75					80	
Thr	Ser	Ser	Cys	Cys	Cys	Glu	Ser	Ser	Arg	Met	Arg	Ala	Met	Ile	Gly
			85						90					95	
Trp	Thr	Pro	Cys	Met	Asp	Val	Lys	Phe	Lys	Asn	Ala	Ser	Ser	Leu	Asn
		100						105					110		
Arg	Thr	Ala	Gly	Leu	Ala	Pro	Gly	Cys	Cys	Gly	Gly	Gly	Pro	Gly	Ala
		115					120					125			
Arg	Thr	Ser	Arg	Glu	Pro	Ser	Pro	Pro	Asp	Ala	Ala	Met	Ala	Ala	Gln
	130					135					140				

	Arg	Ala	Arg	Ala	Pro	Ala	Met	Arg	Thr	Arg	Gly	Gly	Asp	Ala	Ala	Leu
	145					150					155					160
	Cys	Ala	Pro	Glu	Asp	Gly	Trp	Val	Lys	Val	His	Pro	Thr	Pro	Gly	Thr
					165					170						175
5	Met	Leu	Phe	Arg	Glu	Ile	Leu	Leu	Gly	Gln	Met	Gly	Tyr	Thr	Glu	Gly
					180					185					190	
	Gln	Gly	Val	Tyr	Asn	Val	Val	Arg	Ser	Ser	Glu	Ala	Ala	Thr	Arg	Gln
					195				200					205		
	Leu	Gln	Ala	Ala	Ile	Phe	His	Ala	Leu	Leu	Asn	Ala	Thr	Tyr	Asp	Leu
10		210						215					220			
	Glu	Glu	Asp	Trp	Arg	Arg	His	Val	Val	Arg	Leu	Gln	Pro	Gln	Arg	Leu
	225					230					235					240
	Val	Arg	Arg	Tyr	Arg	Asn	Ala	Arg	Glu	Gly	Asp	Ile	Ala	Gly	Val	Ala
					245					250					255	
15	Glu	Arg	Val	Phe	Asp	Thr	Trp	Arg	Cys	Thr	Leu	Arg	Thr	Thr	Leu	Leu
					260				265					270		
	Asp	Phe	Ala	His	Gly	Val	Val	Asp	Cys	Phe	Ala	Pro	Gly	Gly	Pro	Ser
					275				280					285		
	Gly	Pro	Thr	Ser	Phe	Pro	Lys	Tyr	Ile	Asp	Trp	Leu	Thr	Cys	Leu	Gly
20		290						295					300			
	Leu	Val	Pro	Ile	Leu	Arg	Lys	Thr	Arg	Glu	Gly	Glu	Ala	Thr	Gln	Arg
	305					310					315					320
	Leu	Gly	Ala	Phe	Leu	Arg	Gln	His	Thr	Leu	Pro	Arg	Gln	Leu	Ala	Thr
					325					330					335	
25	Val	Ala	Gly	Ala	Ala	Glu	Arg	Ala	Gly	Pro	Gly	Leu	Leu	Glu	Leu	Ala
					340				345					350		
	Val	Ala	Phe	Asp	Ser	Thr	Arg	Met	Ala	Glu	Tyr	Asp	Arg	Val	His	Ile
					355				360					365		
	Tyr	Tyr	Asn	His	Arg	Arg	Gly	Glu	Trp	Leu	Val	Arg	Asp	Pro	Val	Ser
30		370						375					380			
	Gly	Gln	Arg	Gly	Glu	Cys	Leu	Val	Leu	Cys	Pro	Pro	Leu	Trp	Thr	Gly
	385					390					395					400
	Asp	Arg	Leu	Val	Phe	Asp	Ser	Pro	Val	Gln	Arg	Leu	Cys	Pro	Glu	Ile
					405					410					415	
35	Val	Ala	Cys	His	Ala	Leu	Arg	Glu	His	Ala	His	Ile	Cys	Arg	Leu	Arg
					420				425					430		
	Asn	Thr	Ala	Ser	Val	Lys	Val	Leu	Leu	Gly	Arg	Lys	Ser	Asp	Ser	Gly
					435				440					445		
	Val	Ala	Gly	Ala	Ala	Arg	Val	Val	Asn	Lys	Ala	Leu	Gly	Glu	Asp	Asp
40		450						455					460			
	Glu	Thr	Lys	Ala	Gly	Ser	Ala	Ala	Ser	Arg	Leu	Val	Arg	Leu	Ile	Ile
	465					470					475					480
	Asn	Met	Lys	Gly	Met	Arg	His	Val	Gly	Asp	Ile	Asn	Asp	Thr	Val	Arg

		485		490		495
	Ala Tyr Leu Asp Glu Ala Gly Gly His Leu Ile Asp Thr Pro Ala Val					
		500		505		510
	Asp His Thr Leu Pro Gly Phe Gly Lys Gly Gly Thr Gly Arg Gly Ser					
5		515		520		525
	Ala Ala Gln Asp Pro Gly Ala Arg Pro Gln Gln Leu Arg Gln Ala Phe					
		530		535		540
	Gln Thr Ala Val Val Asn Asn Ile Asn Gly Met Leu Glu Gly Tyr Ile					
	545		550		555	560
10	Asn Asn Leu Phe Gly Thr Ile Glu Arg Leu Arg Glu Thr Asn Ala Gly					
		565		570		575
	Leu Ala Thr Gln Leu Gln Ala Arg Asp Arg Glu Leu Arg Arg Ala Gln					
		580		585		590
	Ala Gly Ala Leu Glu Arg Glu Gln Arg Ala Ala Asp Arg Ala Ala Gly					
15		595		600		605
	Gly Gly Ala Gly Arg Pro Ala Glu Ala Asp Leu Leu Arg Ala Asp Tyr					
		610		615		620
	Asp Ile Ile Asp Val Ser Lys Ser Met Asp Asp Asp Thr Tyr Val Ala					
	625		630		635	640
20	Asn Ser Phe Gln His Gln Tyr Ile Pro Ala Tyr Gly Gln Asp Leu Glu					
		645		650		655
	Arg Leu Ser Arg Leu Trp Glu His Glu Leu Val Arg Cys Phe Lys Ile					
		660		665		670
	Leu Arg His Arg Asn Asn Gln Gly Gln Glu Thr Ser Ile Ser Tyr Ser					
25		675		680		685
	Ser Gly Ala Ile Ala Ser Phe Val Ala Pro Tyr Phe Glu Tyr Val Leu					
		690		695		700
	Arg Ala Pro Arg Ala Gly Ala Leu Ile Thr Gly Ser Asp Val Ile Leu					
	705		710		715	720
30	Gly Glu Glu Glu Leu Trp Glu Ala Val Phe Lys Lys Thr Arg Leu Gln					
		725		730		735
	Thr Tyr Leu Thr Asp Val Ala Ala Leu Phe Val Ala Asp Val Gln His					
		740		745		750
	Ala Ala Leu Pro Arg Pro Pro Ser Pro Thr Pro Ala Asp Phe Arg Ala					
35		755		760		765
	Ser Asp Arg Gly Gly Ser Arg Ser Arg Thr Arg Thr Arg Ser Arg Ser					
		770		775		780
	Pro Gly Arg Thr Pro Arg Gly Ala Pro Asp Gln Gly Trp Gly Val Glu					
	785		790		795	800
40	Arg Arg Asp Gly Arg Pro His Ala Arg Arg					
		805		810		

(2) INFORMATION FOR SEQ ID NO:298:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 813 amino acids

(B) TYPE: amino acid

5 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:298:

	Met	Val	Leu	Met	Gly	Arg	Leu	Arg	Asn	Ala	Pro	Glu	Ser	Leu	Thr	Tyr
	1			5						10					15	
	Met	Phe	Cys	Ala	Ala	Ile	Arg	Val	Ala	Pro	Val	Thr	Thr	Gln	Ser	Arg
15				20					25					30		
	Thr	Ser	Leu	Arg	Val	Cys	Thr	His	Val	Leu	Phe	Pro	Asp	Pro	Ala	Leu
				35				40					45			
	Pro	Val	Met	Arg	Tyr	Ala	Ala	Asn	Gly	Asn	Ser	Arg	Ser	Gly	Arg	Pro
				50			55					60				
20	Val	Gly	Thr	Ser	Lys	Ala	Ala	Thr	Ser	Arg	Asn	His	Cys	Arg	Arg	Gly
	65					70				75					80	
	Thr	Cys	Val	Thr	Ser	Ser	Cys	Cys	Cys	Glu	Ser	Ser	Arg	Met	Arg	Ala
						85				90				95		
	Met	Ile	Gly	Trp	Thr	Pro	Cys	Met	Asp	Val	Lys	Phe	Lys	Asn	Ala	Ser
25				100					105					110		
	Ser	Leu	Asn	Arg	Thr	Ala	Gly	Leu	Ala	Pro	Gly	Cys	Cys	Gly	Gly	Gly
				115				120					125			
	Pro	Gly	Ala	Arg	Thr	Ser	Arg	Glu	Pro	Ser	Pro	Pro	Asp	Ala	Ala	Met
				130			135					140				
30	Ala	Ala	Gln	Arg	Ala	Arg	Ala	Pro	Ala	Met	Arg	Thr	Arg	Gly	Gly	Asp
	145					150				155					160	
	Ala	Ala	Leu	Cys	Ala	Pro	Glu	Asp	Gly	Trp	Val	Lys	Val	His	Pro	Thr
					165				170					175		
	Pro	Gly	Thr	Met	Leu	Phe	Arg	Glu	Ile	Leu	Leu	Gly	Gln	Met	Gly	Tyr
35				180				185					190			
	Thr	Glu	Gly	Gln	Gly	Val	Tyr	Asn	Val	Val	Arg	Ser	Ser	Glu	Ala	Ala
				195				200					205			
	Thr	Arg	Gln	Leu	Gln	Ala	Ala	Ile	Phe	His	Ala	Leu	Leu	Asn	Ala	Thr
				210			215				220					
40	Tyr	Asp	Leu	Glu	Glu	Asp	Trp	Arg	Arg	His	Val	Val	Arg	Leu	Gln	Pro
	225					230				235				240		
	Gln	Arg	Leu	Val	Arg	Arg	Tyr	Arg	Asn	Ala	Arg	Glu	Gly	Asp	Ile	Ala
					245				250					255		

Gly Val Ala Glu Arg Val Phe Asp Thr Trp Arg Cys Thr Leu Arg Thr  
 260 265 270  
 Thr Leu Leu Asp Phe Ala His Gly Val Val Asp Cys Phe Ala Pro Gly  
 275 280 285  
 5 Gly Pro Ser Gly Pro Thr Ser Phe Pro Lys Tyr Ile Asp Trp Leu Thr  
 290 295 300  
 Cys Leu Gly Leu Val Pro Ile Leu Arg Lys Thr Arg Glu Gly Glu Ala  
 305 310 315 320  
 Thr Gln Arg Leu Gly Ala Phe Leu Arg Gln His Thr Leu Pro Arg Gln  
 10 325 330 335  
 Leu Ala Thr Val Ala Gly Ala Ala Glu Arg Ala Gly Pro Gly Leu Leu  
 340 345 350  
 Glu Leu Ala Val Ala Phe Asp Ser Thr Arg Met Ala Glu Tyr Asp Arg  
 355 360 365  
 15 Val His Ile Tyr Tyr Asn His Arg Arg Gly Glu Trp Leu Val Arg Asp  
 370 375 380  
 Pro Val Ser Gly Gln Arg Gly Glu Cys Leu Val Leu Cys Pro Pro Leu  
 385 390 395 400  
 Trp Thr Gly Asp Arg Leu Val Phe Asp Ser Pro Val Gln Arg Leu Cys  
 20 405 410 415  
 Pro Glu Ile Val Ala Cys His Ala Leu Arg Glu His Ala His Ile Cys  
 420 425 430  
 Arg Leu Arg Asn Thr Ala Ser Val Lys Val Leu Leu Gly Arg Lys Ser  
 435 440 445  
 25 Asp Ser Gly Val Ala Gly Ala Ala Arg Val Val Asn Lys Ala Leu Gly  
 450 455 460  
 Glu Asp Asp Glu Thr Lys Ala Gly Ser Ala Ala Ser Arg Leu Val Arg  
 465 470 475 480  
 Leu Ile Ile Asn Met Lys Gly Met Arg His Val Gly Asp Ile Asn Asp  
 30 485 490 495  
 Thr Val Arg Ala Tyr Leu Asp Glu Ala Gly Gly His Leu Ile Asp Thr  
 500 505 510  
 Pro Ala Val Asp His Thr Leu Pro Gly Phe Gly Lys Gly Gly Thr Gly  
 515 520 525  
 35 Arg Gly Ser Ala Ala Gln Asp Pro Gly Ala Arg Pro Gln Gln Leu Arg  
 530 535 540  
 Gln Ala Phe Gln Thr Ala Val Val Asn Asn Ile Asn Gly Met Leu Glu  
 545 550 555 560  
 Gly Tyr Ile Asn Asn Leu Phe Gly Thr Ile Glu Arg Leu Arg Glu Thr  
 40 565 570 575  
 Asn Ala Gly Leu Ala Thr Gln Leu Gln Ala Arg Asp Arg Glu Leu Arg  
 580 585 590  
 Arg Ala Gln Ala Gly Ala Leu Glu Arg Glu Gln Arg Ala Ala Asp Arg  
 729



	595					600					605						
	Ala	Ala	Gly	Gly	Gly	Ala	Gly	Arg	Pro	Ala	Glu	Ala	Asp	Leu	Leu	Arg	
	610					615					620						
5	Ala	Asp	Tyr	Asp	Ile	Ile	Asp	Val	Ser	Lys	Ser	Met	Asp	Asp	Asp	Thr	
	625					630					635					640	
	Tyr	Val	Ala	Asn	Ser	Phe	Gln	His	Gln	Tyr	Ile	Pro	Ala	Tyr	Gly	Gln	
	645					650					655						
	Asp	Leu	Glu	Arg	Leu	Ser	Arg	Leu	Trp	Glu	His	Glu	Leu	Val	Arg	Cys	
	660					665					670						
10	Phe	Lys	Ile	Leu	Arg	His	Arg	Asn	Asn	Gln	Gly	Gln	Glu	Thr	Ser	Ile	
	675					680					685						
	Ser	Tyr	Ser	Ser	Gly	Ala	Ile	Ala	Ser	Phe	Val	Ala	Pro	Tyr	Phe	Glu	
	690					695					700						
15	Tyr	Val	Leu	Arg	Ala	Pro	Arg	Ala	Gly	Ala	Leu	Ile	Thr	Gly	Ser	Asp	
	705					710					715					720	
	Val	Ile	Leu	Gly	Glu	Glu	Glu	Leu	Trp	Glu	Ala	Val	Phe	Lys	Lys	Thr	
	725					730					735						
	Arg	Leu	Gln	Thr	Tyr	Leu	Thr	Asp	Val	Ala	Ala	Leu	Phe	Val	Ala	Asp	
	740					745					750						
20	Val	Gln	His	Ala	Ala	Leu	Pro	Arg	Pro	Pro	Ser	Pro	Thr	Pro	Ala	Asp	
	755					760					765						
	Phe	Arg	Ala	Ser	Asp	Arg	Gly	Gly	Ser	Arg	Ser	Arg	Thr	Arg	Thr	Arg	
	770					775					780						
25	Ser	Arg	Ser	Pro	Gly	Arg	Thr	Pro	Arg	Gly	Ala	Pro	Asp	Gln	Gly	Trp	
	785					790					795					800	
	Gly	Val	Glu	Arg	Arg	Asp	Gly	Arg	Pro	His	Ala	Arg	Arg				
	805					810											

(2) INFORMATION FOR SEQ ID NO:299:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 470 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

35

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:299:

40

Met Ala Leu Gly Arg Val Gly Leu Ala Val Gly Leu Trp Gly Leu Leu  
1 5 10 15  
Trp Val Gly Val Val Val Val Leu Ala Asn Asp Gly Arg Thr Ile Thr  
730

	20		25		30
	Val Gly Pro Arg Gly Asn Asn Ala Ala Pro Ser Asp Arg Asn Ala Ser				
	35		40		45
	Ala Pro Arg Thr Thr Pro Thr Pro Pro Gln Pro Arg Lys Ala Thr Lys				
5	50		55		60
	Ser Lys Ala Ser Thr Ala Lys Pro Ala Pro Pro Lys Thr Gly Pro				
	65		70		75
	Pro Lys Thr Ser Ser Glu Pro Val Arg Cys Asn Arg His Asp Pro Leu				
		85		90	
					95
10	Ala Arg Tyr Gly Ser Arg Val Gln Ile Arg Cys Arg Phe Pro Asn Ser				
	100		105		110
	Thr Arg Thr Glu Ser Arg Leu Gln Ile Trp Arg Tyr Ala Thr Ala Thr				
	115		120		125
	Asp Ala Glu Ile Gly Thr Ala Pro Ser Leu Glu Glu Val Met Val Asn				
15	130		135		140
	Val Ser Ala Pro Pro Gly Gly Gln Leu Val Tyr Asp Ser Ala Pro Asn				
	145		150		155
	Arg Thr Asp Pro His Val Ile Trp Ala Glu Gly Ala Gly Pro Gly Asp				
		165		170	
					175
20	Arg Lys Val Val Gly Pro Leu Gly Arg Gln Arg Leu Ile Ile Glu Glu				
	180		185		190
	Leu Thr Leu Glu Thr Gln Gly Met Tyr Tyr Trp Val Trp Gly Arg Thr				
	195		200		205
	Asp Arg Pro Ser Ala Tyr Gly Thr Trp Val Arg Val Arg Val Phe Arg				
25	210		215		220
	Pro Pro Ser Leu Thr Ile His Pro His Ala Val Leu Glu Gly Gln Pro				
	225		230		235
	Phe Lys Ala Thr Cys Thr Ala Ala Thr Tyr Tyr Pro Gly Asn Arg Ala				
		245		250	
					255
30	Glu Phe Val Trp Phe Glu Asp Gly Arg Arg Val Phe Asp Pro Ala Gln				
	260		265		270
	Ile His Thr Gln Thr Gln Glu Asn Pro Asp Gly Phe Ser Thr Val Ser				
	275		280		285
	Thr Val Thr Ser Ala Ala Val Gly Gly Gln Gly Pro Pro Arg Thr Phe				
35	290		295		300
	Thr Cys Gln Leu Thr Trp His Arg Asp Ser Val Ser Phe Ser Arg Arg				
	305		310		315
	Asn Ala Ser Gly Thr Ala Ser Val Leu Pro Arg Pro Thr Ile Thr Met				
		325		330	
					335
40	Glu Phe Thr Gly Asp His Ala Val Cys Thr Ala Gly Cys Val Pro Glu				
	340		345		350
	Gly Val Thr Phe Ala Trp Phe Leu Gly Asp Asp Ser Ser Pro Ala Glu				
	355		360		365

Lys Val Ala Val Ala Ser Gln Thr Ser Cys Gly Arg Pro Gly Thr Ala  
 370 375 380  
 Thr Ile Arg Ser Thr Leu Pro Val Ser Tyr Glu Gln Thr Glu Tyr Ile  
 385 390 395 400  
 5 Cys Arg Leu Ala Gly Tyr Pro Asp Gly Ile Pro Val Leu Glu His His  
 405 410 415  
 Gly Ser His Gln Pro Pro Pro Arg Asp Pro Thr Glu Arg Gln Val Ile  
 420 425 430  
 Arg Ala Val Glu Gly Ala Gly Ile Gly Val Ala Val Leu Val Ala Val  
 10 435 440 445  
 Val Leu Ala Gly Thr Ala Val Val Tyr Leu Thr His Ala Ser Ser Val  
 450 455 460  
 Arg Tyr Arg Arg Leu Arg  
 465 470

15

(2) INFORMATION FOR SEQ ID NO:300:

(i) SEQUENCE CHARACTERISTICS:

20

(A) LENGTH: 536 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:300:

Met Gly Ala Gly Val Pro Trp Thr Gly Ile Lys Arg Ala Gly Gly Pro  
 1 5 10 15  
 30 Ile Thr Val Arg Val Leu Gly Trp Glu Val Ala Gln Lys Ala Thr His  
 20 25 30  
 Pro Cys Cys Ser Cys Pro Arg Glu Ala Val Val Ser Gly Asn Pro Pro  
 35 40 45  
 Arg Cys Ala Gly Arg Ala His Arg Ser Phe Ala Gly Ala Gly Ala Leu  
 35 50 55 60  
 Leu Val Met Ala Leu Gly Arg Val Gly Leu Ala Val Gly Leu Trp Gly  
 65 70 75 80  
 Leu Leu Trp Val Gly Val Val Val Val Leu Ala Asn Asp Gly Arg Thr  
 85 90 95  
 40 Ile Thr Val Gly Pro Arg Gly Asn Asn Ala Ala Pro Ser Asp Arg Asn  
 100 105 110  
 Ala Ser Ala Pro Arg Thr Thr Pro Thr Pro Pro Gln Pro Arg Lys Ala  
 115 120 125

732

Thr Lys Ser Lys Ala Ser Thr Ala Lys Pro Ala Pro Pro Pro Lys Thr  
 130 135 140  
 Gly Pro Pro Lys Thr Ser Ser Glu Pro Val Arg Cys Asn Arg His Asp  
 145 150 155 160  
 5 Pro Leu Ala Arg Tyr Gly Ser Arg Val Gln Ile Arg Cys Arg Phe Pro  
 165 170 175  
 Asn Ser Thr Arg Thr Glu Ser Arg Leu Gln Ile Trp Arg Tyr Ala Thr  
 180 185 190  
 Ala Thr Asp Ala Glu Ile Gly Thr Ala Pro Ser Leu Glu Glu Val Met  
 10 195 200 205  
 Val Asn Val Ser Ala Pro Pro Gly Gly Gln Leu Val Tyr Asp Ser Ala  
 210 215 220  
 Pro Asn Arg Thr Asp Pro His Val Ile Trp Ala Glu Gly Ala Gly Pro  
 225 230 235 240  
 15 Gly Asp Arg Lys Val Val Gly Pro Leu Gly Arg Gln Arg Leu Ile Ile  
 245 250 255  
 Glu Glu Leu Thr Leu Glu Thr Gln Gly Met Tyr Tyr Trp Val Trp Gly  
 260 265 270  
 Arg Thr Asp Arg Pro Ser Ala Tyr Gly Thr Trp Val Arg Val Arg Val  
 20 275 280 285  
 Phe Arg Pro Pro Ser Leu Thr Ile His Pro His Ala Val Leu Glu Gly  
 290 295 300  
 Gln Pro Phe Lys Ala Thr Cys Thr Ala Ala Thr Tyr Tyr Pro Gly Asn  
 305 310 315 320  
 25 Arg Ala Glu Phe Val Trp Phe Glu Asp Gly Arg Arg Val Phe Asp Pro  
 325 330 335  
 Ala Gln Ile His Thr Gln Thr Gln Glu Asn Pro Asp Gly Phe Ser Thr  
 340 345 350  
 Val Ser Thr Val Thr Ser Ala Ala Val Gly Gly Gln Gly Pro Pro Arg  
 30 355 360 365  
 Thr Phe Thr Cys Gln Leu Thr Trp His Arg Asp Ser Val Ser Phe Ser  
 370 375 380  
 Arg Arg Asn Ala Ser Gly Thr Ala Ser Val Leu Pro Arg Pro Thr Ile  
 385 390 395 400  
 35 Thr Met Glu Phe Thr Gly Asp His Ala Val Cys Thr Ala Gly Cys Val  
 405 410 415  
 Pro Glu Gly Val Thr Phe Ala Trp Phe Leu Gly Asp Asp Ser Ser Pro  
 420 425 430  
 Ala Glu Lys Val Ala Val Ala Ser Gln Thr Ser Cys Gly Arg Pro Gly  
 40 435 440 445  
 Thr Ala Thr Ile Arg Ser Thr Leu Pro Val Ser Tyr Glu Gln Thr Glu  
 450 455 460  
 Tyr Ile Cys Arg Leu Ala Gly Tyr Pro Asp Gly Ile Pro Val Leu Glu

465                      470                      475                      480  
 His His Gly Ser His Gln Pro Pro Pro Arg Asp Pro Thr Glu Arg Gln  
                                  485                      490                      495  
 Val Ile Arg Ala Val Glu Gly Ala Gly Ile Gly Val Ala Val Leu Val  
 5                      500                      505                      510  
 Ala Val Val Leu Ala Gly Thr Ala Val Val Tyr Leu Thr His Ala Ser  
                                  515                      520                      525  
 Ser Val Arg Tyr Arg Arg Leu Arg  
                                  530                      535

10

## (2) INFORMATION FOR SEQ ID NO:301:

## (i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 545 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:301:

Met Ser Val Leu Gly Asp Ala Arg His Pro Arg Arg Phe Pro Ser Arg  
   1                      5                      10                      15  
 25 Gly Pro Arg Pro Phe Ser Val Ala Gly Pro Gly Ser Leu Pro Pro Ser  
                                  20                      25                      30  
 Pro Pro Pro Gly Ala Arg Ala Arg Leu Ile Arg Leu Ser Arg Ser Leu  
                                  35                      40                      45  
 Phe Pro Asp Pro Thr Ala Pro Met Asp Leu Leu Val Asp Asp Leu Phe  
 30                      50                      55                      60  
 Ala Asp Ala Asp Gly Val Ser Pro Pro Pro Pro Arg Pro Ala Gly Gly  
   65                      70                      75                      80  
 Pro Lys Asn Thr Pro Ala Ala Pro Pro Leu Tyr Ala Thr Gly Arg Leu  
                                  85                      90                      95  
 35 Ser Gln Ala Gln Leu Met Pro Ser Pro Pro Met Pro Val Pro Pro Ala  
                                  100                      105                      110  
 Ala Leu Phe Asn Arg Leu Leu Asp Asp Leu Gly Phe Ser Ala Gly Pro  
                                  115                      120                      125  
 Ala Leu Cys Thr Met Leu Asp Thr Trp Asn Glu Asp Leu Phe Ser Gly  
 40                      130                      135                      140  
 Phe Pro Thr Asn Ala Asp Met Tyr Arg Glu Cys Lys Phe Leu Ser Thr  
   145                      150                      155                      160  
 Leu Pro Ser Asp Val Ile Asp Trp Gly Asp Ala His Val Pro Glu Arg

		165		170		175										
	Ser	Pro	Ile	Asp	Ile	Arg	Ala	His	Gly	Asp	Val	Ala	Phe	Pro	Thr	Leu
			180						185					190		
	Pro	Ala	Thr	Arg	Asp	Glu	Leu	Pro	Ser	Tyr	Tyr	Glu	Ala	Met	Ala	Gln
5			195					200					205			
	Phe	Phe	Arg	Gly	Glu	Leu	Arg	Ala	Arg	Glu	Glu	Ser	Tyr	Arg	Thr	Val
		210				215						220				
	Leu	Ala	Asn	Phe	Cys	Ser	Ala	Leu	Tyr	Arg	Tyr	Leu	Arg	Ala	Ser	Val
	225				230					235				240		
10	Arg	Gln	Leu	His	Arg	Gln	Ala	His	Met	Arg	Gly	Arg	Asn	Arg	Asp	Leu
				245						250				255		
	Arg	Glu	Met	Leu	Arg	Thr	Thr	Ile	Ala	Asp	Arg	Tyr	Tyr	Arg	Glu	Thr
			260					265						270		
	Ala	Arg	Leu	Ala	Arg	Val	Leu	Phe	Leu	His	Leu	Tyr	Leu	Phe	Leu	Ser
15			275					280					285			
	Arg	Glu	Ile	Leu	Trp	Ala	Ala	Tyr	Ala	Glu	Gln	Met	Met	Arg	Pro	Asp
		290				295						300				
	Leu	Phe	Asp	Gly	Leu	Cys	Cys	Asp	Leu	Glu	Ser	Trp	Arg	Gln	Leu	Ala
	305				310					315				320		
20	Cys	Leu	Phe	Gln	Pro	Leu	Met	Phe	Ile	Asn	Gly	Ser	Leu	Thr	Val	Arg
				325						330				335		
	Gly	Val	Pro	Val	Glu	Ala	Arg	Arg	Leu	Arg	Glu	Leu	Asn	His	Ile	Arg
			340					345						350		
	Glu	His	Leu	Asn	Leu	Pro	Leu	Val	Arg	Ser	Ala	Ala	Ala	Glu	Glu	Pro
25			355					360					365			
	Gly	Ala	Pro	Leu	Thr	Thr	Pro	Pro	Val	Leu	Gln	Gly	Asn	Gln	Ala	Arg
		370				375						380				
	Ser	Ser	Gly	Tyr	Phe	Met	Leu	Leu	Ile	Arg	Ala	Lys	Leu	Asp	Ser	Tyr
	385				390					395				400		
30	Ser	Ser	Val	Ala	Thr	Ser	Glu	Gly	Glu	Ser	Val	Met	Arg	Glu	His	Ala
				405						410				415		
	Tyr	Ser	Arg	Gly	Arg	Thr	Arg	Asn	Asn	Tyr	Gly	Ser	Thr	Ile	Glu	Gly
			420					425						430		
	Leu	Leu	Asp	Leu	Pro	Asp	Asp	Asp	Asp	Ala	Pro	Ala	Glu	Ala	Gly	Leu
35			435					440					445			
	Val	Ala	Pro	Arg	Met	Ser	Phe	Leu	Ser	Ala	Gly	Gln	Arg	Pro	Arg	Arg
		450				455						460				
	Leu	Ser	Thr	Thr	Ala	Pro	Ile	Thr	Asp	Val	Ser	Leu	Gly	Asp	Glu	Leu
	465				470					475				480		
40	Arg	Leu	Asp	Gly	Glu	Glu	Val	Asp	Met	Thr	Pro	Ala	Asp	Ala	Leu	Asp
			485						490				495			
	Asp	Phe	Asp	Leu	Glu	Met	Leu	Gly	Asp	Val	Glu	Ser	Pro	Ser	Pro	Gly
			500					505					510			
								735								

Met Thr His Asp Pro Val Ser Tyr Gly Ala Leu Asp Val Asp Asp Phe  
                   515                  520                  525  
 Glu Phe Glu Gln Met Phe Thr Asp Ala Met Gly Ile Asp Asp Phe Gly  
                   530                  535                  540  
 5 Gly  
       545

## (2) INFORMATION FOR SEQ ID NO:302:

- 10 (i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 490 amino acids  
       (B) TYPE: amino acid  
       (C) STRANDEDNESS: single  
       (D) TOPOLOGY: linear

- 15 (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:302:

20 Met Asp Leu Leu Val Asp Asp Leu Phe Ala Asp Ala Asp Gly Val Ser  
       1                  5                  10                  15  
       Pro Pro Pro Pro Arg Pro Ala Gly Gly Pro Lys Asn Thr Pro Ala Ala  
                   20                  25                  30  
       Pro Pro Leu Tyr Ala Thr Gly Arg Leu Ser Gln Ala Gln Leu Met Pro  
 25                  35                  40                  45  
       Ser Pro Pro Met Pro Val Pro Pro Ala Ala Leu Phe Asn Arg Leu Leu  
                   50                  55                  60  
       Asp Asp Leu Gly Phe Ser Ala Gly Pro Ala Leu Cys Thr Met Leu Asp  
       65                  70                  75                  80  
 30 Thr Trp Asn Glu Asp Leu Phe Ser Gly Phe Pro Thr Asn Ala Asp Met  
                   85                  90                  95  
       Tyr Arg Glu Cys Lys Phe Leu Ser Thr Leu Pro Ser Asp Val Ile Asp  
                   100                  105                  110  
       Trp Gly Asp Ala His Val Pro Glu Arg Ser Pro Ile Asp Ile Arg Ala  
 35                  115                  120                  125  
       His Gly Asp Val Ala Phe Pro Thr Leu Pro Ala Thr Arg Asp Glu Leu  
                   130                  135                  140  
       Pro Ser Tyr Tyr Glu Ala Met Ala Gln Phe Phe Arg Gly Glu Leu Arg  
       145                  150                  155                  160  
 40 Ala Arg Glu Glu Ser Tyr Arg Thr Val Leu Ala Asn Phe Cys Ser Ala  
                   165                  170                  175  
       Leu Tyr Arg Tyr Leu Arg Ala Ser Val Arg Gln Leu His Arg Gln Ala  
                   180                  185                  190

His Met Arg Gly Arg Asn Arg Asp Leu Arg Glu Met Leu Arg Thr Thr  
 195 200 205  
 Ile Ala Asp Arg Tyr Tyr Arg Glu Thr Ala Arg Leu Ala Arg Val Leu  
 210 215 220  
 5 Phe Leu His Leu Tyr Leu Phe Leu Ser Arg Glu Ile Leu Trp Ala Ala  
 225 230 235 240  
 Tyr Ala Glu Gln Met Met Arg Pro Asp Leu Phe Asp Gly Leu Cys Cys  
 245 250 255  
 Asp Leu Glu Ser Trp Arg Gln Leu Ala Cys Leu Phe Gln Pro Leu Met  
 10 260 265 270  
 Phe Ile Asn Gly Ser Leu Thr Val Arg Gly Val Pro Val Glu Ala Arg  
 275 280 285  
 Arg Leu Arg Glu Leu Asn His Ile Arg Glu His Leu Asn Leu Pro Leu  
 290 295 300  
 15 Val Arg Ser Ala Ala Ala Glu Glu Pro Gly Ala Pro Leu Thr Thr Pro  
 305 310 315 320  
 Pro Val Leu Gln Gly Asn Gln Ala Arg Ser Ser Gly Tyr Phe Met Leu  
 325 330 335  
 Leu Ile Arg Ala Lys Leu Asp Ser Tyr Ser Ser Val Ala Thr Ser Glu  
 20 340 345 350  
 Gly Glu Ser Val Met Arg Glu His Ala Tyr Ser Arg Gly Arg Thr Arg  
 355 360 365  
 Asn Asn Tyr Gly Ser Thr Ile Glu Gly Leu Leu Asp Leu Pro Asp Asp  
 370 375 380  
 25 Asp Asp Ala Pro Ala Glu Ala Gly Leu Val Ala Pro Arg Met Ser Phe  
 385 390 395 400  
 Leu Ser Ala Gly Gln Arg Pro Arg Arg Leu Ser Thr Thr Ala Pro Ile  
 405 410 415  
 Thr Asp Val Ser Leu Gly Asp Glu Leu Arg Leu Asp Gly Glu Glu Val  
 30 420 425 430  
 Asp Met Thr Pro Ala Asp Ala Leu Asp Asp Phe Asp Leu Glu Met Leu  
 435 440 445  
 Gly Asp Val Glu Ser Pro Ser Pro Gly Met Thr His Asp Pro Val Ser  
 450 455 460  
 35 Tyr Gly Ala Leu Asp Val Asp Asp Phe Glu Phe Glu Gln Met Phe Thr  
 465 470 475 480  
 Asp Ala Met Gly Ile Asp Asp Phe Gly Gly  
 485 490

40

(2) INFORMATION FOR SEQ ID NO:303:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 552 amino acids



(B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:303:

10	Met	Arg	Gly	Gly	Gly	Arg	Glu	Met	Ser	Val	Leu	Gly	Asp	Ala	Arg	His
	1				5					10					15	
	Pro	Arg	Arg	Phe	Pro	Ser	Arg	Gly	Pro	Arg	Pro	Phe	Ser	Val	Ala	Gly
				20					25					30		
	Pro	Gly	Ser	Leu	Pro	Pro	Ser	Pro	Pro	Pro	Gly	Ala	Arg	Ala	Arg	Leu
				35					40					45		
15	Ile	Arg	Leu	Ser	Arg	Ser	Leu	Phe	Pro	Asp	Pro	Thr	Ala	Pro	Met	Asp
				50				55					60			
	Leu	Leu	Val	Asp	Asp	Leu	Phe	Ala	Asp	Ala	Asp	Gly	Val	Ser	Pro	Pro
	65					70					75					80
	Pro	Pro	Arg	Pro	Ala	Gly	Gly	Pro	Lys	Asn	Thr	Pro	Ala	Ala	Pro	Pro
20						85				90					95	
	Leu	Tyr	Ala	Thr	Gly	Arg	Leu	Ser	Gln	Ala	Gln	Leu	Met	Pro	Ser	Pro
					100					105					110	
	Pro	Met	Pro	Val	Pro	Pro	Ala	Ala	Leu	Phe	Asn	Arg	Leu	Leu	Asp	Asp
				115					120						125	
25	Leu	Gly	Phe	Ser	Ala	Gly	Pro	Ala	Leu	Cys	Thr	Met	Leu	Asp	Thr	Trp
				130				135						140		
	Asn	Glu	Asp	Leu	Phe	Ser	Gly	Phe	Pro	Thr	Asn	Ala	Asp	Met	Tyr	Arg
	145					150					155				160	
	Glu	Cys	Lys	Phe	Leu	Ser	Thr	Leu	Pro	Ser	Asp	Val	Ile	Asp	Trp	Gly
30					165					170				175		
	Asp	Ala	His	Val	Pro	Glu	Arg	Ser	Pro	Ile	Asp	Ile	Arg	Ala	His	Gly
				180					185					190		
	Asp	Val	Ala	Phe	Pro	Thr	Leu	Pro	Ala	Thr	Arg	Asp	Glu	Leu	Pro	Ser
				195				200					205			
35	Tyr	Tyr	Glu	Ala	Met	Ala	Gln	Phe	Phe	Arg	Gly	Glu	Leu	Arg	Ala	Arg
				210				215					220			
	Glu	Glu	Ser	Tyr	Arg	Thr	Val	Leu	Ala	Asn	Phe	Cys	Ser	Ala	Leu	Tyr
	225					230					235				240	
	Arg	Tyr	Leu	Arg	Ala	Ser	Val	Arg	Gln	Leu	His	Arg	Gln	Ala	His	Met
40					245					250					255	
	Arg	Gly	Arg	Asn	Arg	Asp	Leu	Arg	Glu	Met	Leu	Arg	Thr	Thr	Ile	Ala
				260					265					270		
	Asp	Arg	Tyr	Tyr	Arg	Glu	Thr	Ala	Arg	Leu	Ala	Arg	Val	Leu	Phe	Leu

	275		280		285
	His Leu Tyr Leu Phe Leu Ser Arg Glu Ile Leu Trp Ala Ala Tyr Ala				
	290		295		300
5	Glu Gln Met Met Arg Pro Asp Leu Phe Asp Gly Leu Cys Cys Asp Leu				
	305		310		315 320
	Glu Ser Trp Arg Gln Leu Ala Cys Leu Phe Gln Pro Leu Met Phe Ile				
		325		330	335
	Asn Gly Ser Leu Thr Val Arg Gly Val Pro Val Glu Ala Arg Arg Leu				
		340		345	350
10	Arg Glu Leu Asn His Ile Arg Glu His Leu Asn Leu Pro Leu Val Arg				
		355		360	365
	Ser Ala Ala Ala Glu Glu Pro Gly Ala Pro Leu Thr Thr Pro Pro Val				
		370		375	380
	Leu Gln Gly Asn Gln Ala Arg Ser Ser Gly Tyr Phe Met Leu Leu Ile				
15	385		390		395 400
	Arg Ala Lys Leu Asp Ser Tyr Ser Ser Val Ala Thr Ser Glu Gly Glu				
		405		410	415
	Ser Val Met Arg Glu His Ala Tyr Ser Arg Gly Arg Thr Arg Asn Asn				
		420		425	430
20	Tyr Gly Ser Thr Ile Glu Gly Leu Leu Asp Leu Pro Asp Asp Asp Asp				
		435		440	445
	Ala Pro Ala Glu Ala Gly Leu Val Ala Pro Arg Met Ser Phe Leu Ser				
		450		455	460
	Ala Gly Gln Arg Pro Arg Arg Leu Ser Thr Thr Ala Pro Ile Thr Asp				
25	465		470		475 480
	Val Ser Leu Gly Asp Glu Leu Arg Leu Asp Gly Glu Glu Val Asp Met				
		485		490	495
	Thr Pro Ala Asp Ala Leu Asp Asp Phe Asp Leu Glu Met Leu Gly Asp				
		500		505	510
30	Val Glu Ser Pro Ser Pro Gly Met Thr His Asp Pro Val Ser Tyr Gly				
		515		520	525
	Ala Leu Asp Val Asp Asp Phe Glu Phe Glu Gln Met Phe Thr Asp Ala				
		530		535	540
	Met Gly Ile Asp Asp Phe Gly Gly				
35	545		550		

**What is claimed is:**

1. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:
  - (a) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising an amino acid sequence of Table 1, 2, 3 or 4;
  - (b) a polynucleotide having at least a 70% identity to a polynucleotide encoding a mature polypeptide expressed by the gene contained in the HSV-2 of deposited strain VR-2546 that was sequenced to obtain a polynucleotide sequence of Table 1, 2 or 3;
  - (c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 70% identical to an amino acid sequence of Table 1, 2, 3 or 4;
  - (d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and
  - (e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).
2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.
3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.
4. The polynucleotide of Claim 2 comprising the nucleic acid sequence selected from the group consisting of the nucleic acid sequences set forth in Table 1, 2 and 3.
5. The polynucleotide of Claim 2 which encodes a polypeptide comprising an amino acid sequence selected from the group consisting of the amino acid sequences set forth in Table 1, 2, 3 and 4.
6. A vector comprising the polynucleotide of Claim 1.
7. A host cell comprising the vector of Claim 6.
8. A process for producing a polypeptide comprising expressing in the host cell of Claim 7 a polypeptide encoded by said polynucleotide.
9. A process for producing a polypeptide or fragment thereof comprising culturing a host cell of Claim 7 under conditions sufficient for the production of said polypeptide or fragment.
10. A polypeptide comprising an amino acid sequence which is at least 70% identical to an amino acid sequence selected from the group consisting of the amino acid sequences or fragments thereof set forth in Table 1, 2, 3 and 4.
11. A polypeptide comprising an amino acid sequence selected from the group consisting of the amino acid sequences or fragments thereof set forth in Table 1, 2, 3, and 4.
12. An antibody generated against the polypeptide of claim 10.

13. An antagonist or agonist of the activity or expression of the polypeptide of claim 10.
14. A method for the treatment or prevention of disease of an individual comprising administering to the individual a therapeutically effective amount of the polypeptide of claim 10.
- 5 15. A method for the treatment of an individual having medical need to inhibit a viral polypeptide comprising administering to the individual a therapeutically effective amount of the antagonist of Claim 13.
16. A process for diagnosing a disease related to expression or activity of the polypeptide of claim 10 in an individual comprising
- 10 (a) determining a nucleic acid sequence encoding said polypeptide, and/or
- (b) analyzing for the presence or amount of said polypeptide in a sample derived from the individual.
17. A method for identifying compounds which inhibit or activate the polypeptide of claim 10 comprising
- 15 (a) contacting a composition comprising the polypeptide with the compound to be screened under conditions to permit interaction between the compound and the polypeptide to assess the interaction of a compound, such interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the polypeptide with the compound; and
- 20 (b) determining whether the compound activates or inhibits polypeptide by detecting the presence or absence of the signal generated from the interaction of the compound with the polypeptide.
18. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with the polypeptide of Claim 10, or a variant thereof,
- 25 adequate to produce antibody and/or T cell immune response to protect said animal from disease.
19. A method of inducing immunological response in a mammal which comprises delivering a nucleic acid vector to direct expression of a polypeptide of Claim 10, or a variant thereof, for expressing said polypeptide *in vivo* in order to induce an immunological
- 30 response to produce antibody and/ or T cell immune response to protect said animal from disease.
20. The isolated polynucleotide of claim 1 wherein said nucleotide is selected from the group consisting of:
- (a) a polynucleotide having at least a 90% identity to a polynucleotide encoding a
- 35 polypeptide comprising the amino acid sequence of Table 1, 2, 3 or 4;

- (b) a polynucleotide having at least a 90% identity to a polynucleotide encoding the same mature polypeptide expressed by the gene contained in the HSV-2 of the deposited strain VR-2546 that was sequenced to obtain a polynucleotide sequence of Table 1, 2 or 3;
- (c) a polynucleotide encoding a polypeptide comprising an amino acid sequence  
5 which is at least 90% identical to the amino acid sequence of Table 1, 2, 3 or 4;
- (d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and
- (e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).
- 10 21. The isolated polynucleotide of Claim 1 selected from the group consisting of
- (a) a polynucleotide having at least a 95% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1, 2, 3 or 4;
- (b) a polynucleotide having at least a 95% identity to a polynucleotide encoding the same mature polypeptide expressed by the gene contained in the HSV-2 of the deposited strain  
15 VR-2546 that was sequenced to obtain a polynucleotide sequence of Table 1, 2 or 3;
- (c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 95% identical to the amino acid sequence of Table 1, 2, 3 or 4;
- (d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and
- 20 (e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).
22. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:
- (a) a polynucleotide having at least a 50% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1, 2, 3 or 4 and obtained from a  
25 prokaryotic species other than HSV-2;
- (b) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 50% identical to the amino acid sequence of Table 1, 2, 3 or 4 and obtained from a prokaryotic species other than HSV-2; and
- 30 (c) a polynucleotide which is complementary to the polynucleotide of (a) or (b).
23. An isolated polypeptide having one of the amino acid sequences given in Table 1, 2, 3 or 4.
24. An isolated nucleic acid encoding one of the amino acid sequences of Claim 1 and nucleic acid sequences capable of hybridizing therewith under stringent  
35 conditions.

25. A recombinant vector comprising the nucleic acid sequences of Claim 24 and host cells transformed or transfected therewith.
26. A method of identifying an antiviral compound comprising contacting candidate compounds with a polypeptide of Claim 10 and selecting those compounds  
5 capable of inhibiting the bioactivity of said polypeptide.
27. Antiviral compounds identified by the method of Claim 26.
28. An isolated polypeptide having an amino acid sequence or fragment thereof given in Table 1, 2, 3 or 4.
29. An isolated nucleic acid encoding one of the amino acid sequences of  
10 Claim 28 and nucleic acid sequences capable of hybridizing therewith under stringent conditions.
30. A method of identifying an antiviral compound comprising contacting candidate compounds with a polypeptide of Claim 28 and selecting those compounds capable of inhibiting the bioactivity of said polypeptide.

15

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/20016

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C07H 17/00; C12N 15/00, 5/00

US CL : 424/199.1, 204.1, 229.1

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/199.1, 204.1, 229.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, CAS ONLINE, CAPLUS, MEDLINE

search terms: HERPESVIRUS TYPE 2, HSV-2, NUCLEOTIDE?, PROTEIN OR POLYPEPTIDE

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 95/06055 A1 (SMITHKLINE BEECHAM CORPORATION) 02 March 1995, page 1, abstract.	1-11, 20-25, 28, and 291-
Y	WO 95/16779 A1 (SMITHKLINE BEECHAM CORPORATION) 22 June 1995, see sequence listings on pages 16-21.	1-11, 20-25, 28, and 29
Y	STEFFY et al. Nucleotide sequence of herpes simplex virus type 2 gene encoding the protease and capsid protein ICP35. Journal of General Virology. 1995, Vol. 76, pages 1069-1072, see abstract.	1-11, 20-25, 28, and 29



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y* document of particular relevance, the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

11 FEBRUARY 1998

Date of mailing of the international search report

25 MAR 1998

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

ALI R. SALIMI

Telephone No. (703) 308-0196

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/20016

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:  
1-11, 20-25, 28, and 29
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐  
☒

- The additional search fees were accompanied by the applicant's protest.  
No protest accompanied the payment of additional search fees.



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/20016

### BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-9, 20-22, 24-25, and 29, drawn to Polynucleotide encoding HSV-2 and its first use (making protein).

Group II, claim(s) 10, 11, 23, and 28, drawn to polypeptide(s) and their second use (diagnosis, second product).

Group III, claim(s) 12, and 14, drawn to antibodies (third product).

Group IV, claim(s) 13, and 15 drawn to antagonist or agonist (fourth product).

Group V, claim(s) 16, 17, 26, 27, and 30, drawn to method of diagnosing and identifying polypeptide (fifth product).

Group VI, claim(s) 18, and 19, drawn to vaccine (sixth product).

and it considers that the International Application does not comply with the requirements of unity of invention (Rules 13.1, 13.2 and 13.3) for the reasons indicated below:

The inventions listed as Groups I-VI do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups I-VI are directed to isolation and use of nucleic acid of herpes simplex virus type 2 (HSV-2). They are further directed to various percent identity of the said virus nucleotide. In addition they are directed to expression and translation of the isolated nucleic acids of claim 1 in various diagnostic and pharmaceutical compositions. The claims are linked by the disclosed nucleic acid sequences of HSV-2 of claim 1. However, such does not constitute a special technical feature because an isolated nucleotide sequence of herpes virus encoding capsid and protease have been reported previously Steffy et al (Journal of General Virology, Vol. 76, 1995) and also Smithkline Beecham Corporation (WO 95/06055). The cited evidences prove that the technical feature of Group I, disclosed nucleic acid sequences of HSV-2, does not make a contribution over the prior art. The claims are not so linked by a special technical feature within the meaning of PCT Rule 13.2 so as to form a single inventive concept, accordingly, the unity of the invention is lacking among all groups.